

London School of Economics and Political Science

***Multiple Criteria Decision Analysis for Assessing the Value of
New Medical Technologies***

*Researching, developing and applying a new value framework for the purpose
of Health Technology Assessment*

Aris Nikolaos Angelis

A thesis submitted to the Department of Social Policy of the London
School of Economics for the degree of Doctor of Philosophy, London,
August, 2017

Declaration of Authorship

I certify that the thesis I have presented for examination for the Social Policy PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it – see ‘Statement of conjoint work’).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 65,750 words (Chapters 1 – 8), including footnotes, excluding references and appendices.

Statement of Conjoint Work

Part of the work presented in Chapters 3-7 of this thesis has been published or is currently under review in peer review academic journals¹, co-authored with Prof Panos Kanavos from London School of Economics (London, UK), Dr Ansgar Lange from London School of Economics (London, UK), Prof Gilberto Montibeller from Loughborough University (Loughborough, UK) and Dr Daniel Hochhauser from University College London Hospital (London, UK). For the published aspects of this work, I conceived the studies, designed the studies, collected the data, analysed the data, interpreted the findings and drafted the papers. I confirm that I am fully responsible for the entirety of the work presented in this doctoral thesis other than where I have cited the relevant work of others.

Chapter 3 has been published with Prof Panos Kanavos as co-author: Angelis A, Kanavos P (2016). Value-Based Assessment of New Medical Technologies: Towards A Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment. *PharmacoEconomics* May; 34(5):435-46 (Supplied in *Appendix* to Published Articles). I devised the study, wrote the first draft, contributed to subsequent drafts, and am the guarantor. Panos Kanavos supervised the study, provided methodological guidance, and contributed to the published manuscript version.

Chapter 4 has been published with Dr Ansgar Lange and Prof Panos Kanavos as co-authors: Angelis A, Lange A, Kanavos P (2017). Using health technology assessment to assess the value of new medicines: Results of a systematic review and expert consultation across eight European countries. *European Journal of Health Economics*. In press; DOI: 10.1007/s10198-017-0871-0. I devised the study, collected the data, analysed the data, wrote the first draft, contributed to subsequent drafts, and am the guarantor. Ansgar Lange contributed to the published manuscript version. Panos Kanavos supervised the study, provided methodological guidance, and contributed to the published manuscript version.

Chapter 5 has been published with Prof Panos Kanavos as co-author: Angelis A, Kanavos P (2017). Multiple Criteria Decision Analysis (MCDA) for Evaluating New Medicines in Health Technology Assessment and Beyond: the Advance Value Framework. *Social Science*

¹ Chapter 7 has been submitted for peer review and it is single-authored by myself as: Angelis A. Applying Multiple Criteria Decision Analysis (MCDA) in the context of Health Technology Assessment (HTA): a case study on metastatic prostate cancer with the Swedish HTA agency TLV (Under Review at Value in Health)

& *Medicine*. In press; doi.org/10.1016/j.socscimed.2017.06.024. I devised the study, collected the data, analysed the data, wrote the first draft, contributed to subsequent drafts, and am the guarantor. Panos Kanavos supervised the study, provided methodological guidance, and contributed to the published manuscript version.

Chapter 6 has been submitted for peer review with Prof Gilberto Montibeller, Dr Daniel Hochhauser and Prof Panos Kanavos as co-authors: Angelis A, Montibeller G, Hochhauser D, Kanavos P. Multiple Criteria Decision Analysis (MCDA) in the Context of Health Technology Assessment: a Simulation Exercise on Metastatic Colorectal Cancer with Multiple Stakeholders in the English Settings (Under Review at *BMC Medical Informatics and Decision Making*). I devised the study, collected the data, designed the model, run the model, analysed the data, wrote the first draft, contributed to subsequent drafts, and am the guarantor. Gilberto Montibeller provided methodological suggestions, facilitated the workshop and contributed to the submitted manuscript version. Daniel Hochhauser provided clinical insights and contributed to the submitted manuscript version. Panos Kanavos supervised the study, provided methodological guidance, and contributed to the submitted manuscript version.

Other Relevant Work

During my PhD, I co-authored a number of peer-reviewed publications, which, although highly relevant to the work presented here, did not directly contribute to the empirical chapters of my thesis. These are listed below:

Angelis A, Kanavos P (2016). A Critique to ASCO's Value Assessment Framework for Cancer Treatments: Putting Methodological Robustness First. *Journal of Clinical Oncology*. August; 34(24): 2935-2936 (Supplied in *Appendix* to Published Articles)

Angelis A, Kanavos P, Montibeller G (2016). Resource Allocation and Priority Setting in Healthcare: A Multi-criteria Decision Analysis Problem of Value? *Global Policy*. In press; DOI: 10.1111/1758-5899.12387 (Supplied in *Appendix* to Published Articles)

Tervonen T, Naci H, van Valkenhoef G, Ades AE, Angelis A, Hillege HL, Postmus D (2015). Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing Among Statins in Primary Prevention. *Medical Decision Making* 35 (7). 859-871. ISSN 1552-681X (Supplied in *Appendix* to Published Articles)

Kanavos P, Nicod E, Visintin E, Angelis A (2016). Valuing Value: a Retrospective and Prospective Assessment of New Medical Technologies (Under Review at *Health Policy*)

Acknowledgements

I would like to acknowledge that the research activities of the thesis have been carried out under the auspices of Advance-HTA, an FP-7 research project that has received funding from the European Commission, DG Research (Grant agreement no: 305983) and led by the London School of Economics (LSE). The views represented in this thesis do not necessarily reflect the views of the European Commission. Parts of this thesis have been submitted to the European Commission as deliverable of Work Package 2 of the Advance-HTA project.

I would like to express my gratitude to my PhD supervisor Prof Panos Kanavos (LSE) for providing me with the intellectual freedom and academic support to pursue my research interests and for his enthusiasm, encouragement, guidance, and feedback throughout the four years of my PhD. I am also thankful to my second PhD supervisor Prof Alistair McGuire (LSE) for his valuable feedback throughout the years.

I would like to specially thank Dr Matteo Galizzi (LSE) and Dr Monica Oliveira (University of Lisbon) for acting as my PhD viva examiners, thorough reviewing my thesis and providing me with rigorous feedback and ways of improving my work.

I am particularly grateful to Prof Gilberto Montibeller (Loughborough University) for teaching me decision analysis skills and providing me with fundamental theoretical direction; I would like to acknowledge his input in Chapters 4 and 6 where he provided valuable feedback in earlier versions. I would like to offer my special thanks to Prof Larry Phillips (LSE) and Prof Carlos Bana e Costa (University of Lisbon) for sharing with me their long expertise in the decision sciences field; I would like to acknowledge their valuable and constructive feedback for Chapters 6 and 7. I would also like to thank the partners of Advance-HTA consortium for their contributions in Work Package 2, including their views and opinions during the meetings of the consortium; I would like to acknowledge helpful comments and suggestions for Chapters 4, and 5.

I would like to thank Dr Sotiris VANDOROS (King's College London), Dr Huseyin Naci (LSE), Dr Elena Nicod (LSE), Alesandra Ferrario (LSE) and Olivier Wouters (LSE) for discussing with me research matters, reviewing parts of my thesis, and providing helpful comments and suggestions. I would also like to acknowledge and thank my colleagues Olina Efthymiadou and Erica Vistin at the Medical Technology Research Group (LSE) for working with me and offering valuable research assistance in various aspects of evidence collection concerning the performance of the alternative treatments assessed as part of Chapters 6 and 7.

I am grateful to the following experts for reviewing the results of Chapter 4 (systematic literature review) and providing critical feedback that shaped the content of the Chapter: Karl Arnberg (TLV), Karen Brigham (University Paris Est), Agnese Cangini (AIFA), Guillaume Dedet (World Health Organisation), Jaime Espin (Andalusian School of Public Health), Mathilde Grande (HAS), Kinga Grzywacz (AOTMiT), Patrick Jeurissen (Dutch Ministry of Health, Welfare and Sport & Radboud University Nijmegen Medical Centre), Douglas Lundin (TLV), Armando Magrelli (National Centre for Rare Diseases, Istituto Superiore di Sanità), Dimitra Panteli (Berlin University of Technology), Dominique Polton (National Health Insurance Fund), Pierluigi Russo (AIFA), Annalisa Sammarco (AIFA), Pedro Serrano (Canary Islands Foundation for Research and Health), Giovanni Tafuri (AIFA & EMA) and Anna Zawada (AOTMiT).

I would like to acknowledge the feedback that has been received from the participants at the HTAi conference in Oslo (June 2015), the YoungOR19 conference of the OR Society in Birmingham (September 2015) and the PPRI conference of the WHO Collaborating Centre for Pharmaceutical Pricing and Reimbursement Policies in Vienna (October 2015); I am grateful for all comments and suggestions received. I would also like to thank the participants of the Advance-HTA dissemination and capacity building workshops in Mexico City (November 2014), Santiago (September 2015), and Warsaw (September 2014 and 2015), for providing constructive comments and their perspectives on the development of the value tree and the value framework, which helped shaping the content of Chapter 6.

I would like to thank the participants of the decision conference that attended a one-day workshop at London School of Economics (LSE) and whose preferences were elicited acting as the basis of the analysis for Chapter 6: Keith Abrams (University of Leicester), Stuart Barber (Beating Bowel Cancer), Andres Freiberg (NICE International), Rob Glynne-Jones (Mount Vernon Hospital), Burrinder Grewal, Daniel Hochhauser (UCL Hospital), Barbara Moss, Elena Nicod (LSE), Mark Ross, Caroline Rudisill (LSE), David Taylor (UCL), Sotiris Vadoros (King's College London), John Wright (LSE). I would like to acknowledge Daniel Hochhauser (UCL Hospital) for his valuable clinical insight and his contribution to the manuscript submitted for publication that formed part of Chapter 6.

I would also like to thank the staff of the Dental and Pharmaceutical Benefits Agency in Sweden (TLV) that participated in two half-day workshops in Stockholm and whose preferences were elicited acting as the basis of the analysis for Chapter 7: Loudin Daoura, Niklas Hedberg, Daniel Högberg, Douglas Lundin. Special thanks go to Karl Arnberg and

Douglas Lundin for helping to organise the workshop. I would like to acknowledge Mark Linch (UCL Hospital) for his valuable clinical insight and his contribution to the manuscript submitted for publication that formed part of Chapter 7.

Special thanks go out to my old friends Dr Kyriakoulis Resvanis (Imperial College London) and Dr Themistocles Resvanis (Massachusetts Institute of Technology) for their help and thoughts in preparation of this manuscript and for the many long and fruitful discussions. Thanks also go out to my friends and flatmates Kaf, Koulis, Jimmy and Grafas for putting up with me and making sure that there was never a mundane moment.

Last but not least I would like to thank my parents and family for their continuous support.

Abstract

Introduction: Current evaluation approaches for new medical technologies are problematic for a plethora of reasons relating to measuring their expected costs and consequences, but also due to hurdles in turning assessed information into coverage decisions. Most adopted

methodologies focus on a limited number of value dimensions, despite the fact that the value of new medicines is multi-dimensional in nature. Explicit elicitation of social value trade-offs is not possible and decision-makers may adopt intuitive or heuristic modes for simplification purposes, based on *ad hoc* procedures that might lead to arbitrary decisions.

Objectives: The objective of the present thesis is to develop and empirically test a methodological framework that can be used to assess the overall value of new medical technologies by explicitly capturing multiple aspects of value while allowing for their trade-offs, through the incorporation of decision-makers' preferences in a structured and transparent way. The research hypothesis is that Multiple Criteria Decision Analysis (MCDA) can provide a methodological option for the evaluation of new medicines in the context of Health Technology Assessment (HTA), to support decision-making and contribute to more efficient resource allocation.

Methods and Empirical Evidence: The first paper proposes a conceptual methodological process, based on multi-attribute value theory (MAVT) methods comprising five distinct phases, outlining the stages involved in each phase and discusses their relevance in the HTA context. The second paper conducts a systematic literature review and expert consultation in order to investigate the practices, processes and policies of value-assessment for new medicines across eight European countries and identifies the evaluation criteria employed and how these inform coverage recommendations as part of HTA. The third paper develops a MAVT value framework for HTA, incorporating a generic value tree for new medicines composed from different levels of criteria that fall under five value domains (i.e. therapeutic, safety, burden of disease, innovation and socio-economic), together with a selection of scoring, weighting and aggregating techniques. In the fourth and fifth papers, the value framework is tested empirically by conducting two real-world case studies: in the first, the value tree is adapted for the evaluation of second-line biological treatments for metastatic colorectal cancer (mCRC) patients having received prior oxaliplatin-based chemotherapy; in the second, the value tree is conditioned for the evaluation of third-line treatments for metastatic castrate resistant prostate cancer (mCRPC) patients having received prior docetaxel chemotherapy. Both case studies were informed by decision conferences with relevant expert panels. In the mCRC decision conference multiple stakeholders participated reflecting the composition of the English National Institute for Health and Care Excellence (NICE) technology appraisal committees, whereas in the mCRPC decision conference a

group of evaluators participated from the Swedish Dental and Pharmaceutical Benefits Agency (TLV), thereby adopting the TLV decision-making perspective.

Policy Implications: The value scores produced from the MCDA process reflect a more comprehensive benefit metric that embeds the preferences of stakeholders and decision-makers across a number of explicit evaluation criteria. The incorporation of alternative treatments' purchasing costs can then be used to derive incremental cost value ratios based on which the treatments can be ranked on 'value-for-money' grounds, reflecting their incremental cost relative to incremental value.

Conclusion: The MCDA value framework developed can aid HTA decision-makers by allowing them to explicitly consider multiple criteria and their relative importance, enabling them to understand and incorporate their own preferences and value trade-offs in a constructed and transparent way. It can be turned into a flexible decision tool for resource allocation purposes in the coverage of new medicines by payers but could also be adapted for other decision-making contexts along their development, regulation and use.

Table of contents

DECLARATION OF AUTHORSHIP	2
STATEMENT OF CONJOINT WORK.....	3
OTHER RELEVANT WORK	5
ACKNOWLEDGEMENTS	6

ABSTRACT	8
LIST OF CHAPTERS.....	13
LIST OF FIGURES.....	14
LIST OF TABLES	15
LIST OF BOXES.....	16
LIST OF ABBREVIATIONS	17
CHAPTER 1 – INTRODUCTION TO HEALTH TECHNOLOGY ASSESSMENT, THESIS RESEARCH QUESTIONS AND OVERVIEW	21
1.1 INTRODUCTION AND THE NEED FOR IMPROVING THE WAY WE MEASURE THE VALUE OF HEALTH CARE INTERVENTIONS	21
1.2 RESEARCH QUESTIONS AND THESIS OVERVIEW	33
1.3 CONCLUSION.....	43
CHAPTER 2 – CURRENT METHODS OF VALUE ASSESSMENT OF MEDICAL TECHNOLOGIES	45
2.1 CURRENT METHODS OF VALUE ASSESSMENT IN HEALTH TECHNOLOGY ASSESSMENT	45
2.2 LIMITATIONS OF CURRENT VALUE ASSESSMENT APPROACHES.....	49
2.3 SUMMARY OF GAPS IN THE LITERATURE AND LIKELY WAYS FORWARD.....	52
2.4 CONCLUSION.....	54
CHAPTER 3 – PAPER 1	56
CONCEPTUAL FRAMEWORK: TOWARDS A ROBUST METHODOLOGICAL FRAMEWORK FOR THE APPLICATION OF MULTIPLE CRITERIA DECISION ANALYSIS IN THE CONTEXT OF HEALTH TECHNOLOGY ASSESSMENT	56
SUMMARY.....	56
3.1 INTRODUCTION.....	56
3.2 A METHODOLOGICAL PROCESS APPLYING MCDA PRINCIPLES IN HTA AND VALUE ASSESSMENT OF MEDICAL TECHNOLOGIES	59
3.3 MODEL-BUILDING AND THE CONSTRUCTION OF A VALUE TREE: PROPERTIES TO ENSURE A ROBUST MCDA MODEL	69
3.4. ‘INCREMENTAL’ VERSUS ‘CLEAN SLATE’ MCDA APPROACHES AND LINK TO POLICY MAKING.....	73
3.5 CONCLUSION.....	75
CHAPTER 4 – PAPER 2	76
EVALUATION CRITERIA AND DECISION-MAKING PRACTICES IN HEALTH TECHNOLOGY ASSESSMENT: A SYSTEMATIC REVIEW AND EXPERT CONSULTATION ACROSS 8 EU COUNTRIES	76
SUMMARY.....	77
4.1 BACKGROUND	77
4.2 METHODS	79

4.3 RESULTS.....	86
4.4 DISCUSSION	117
4.5 CONCLUSION.....	120
CHAPTER 5 – PAPER 3	121
METHODOLOGICAL FRAMEWORK: A MULTIPLE CRITERIA DECISION ANALYSIS VALUE MEASUREMENT MODEL AND TECHNIQUES FOR THE EVALUATION OF NEW MEDICINES IN HEALTH TECHNOLOGY ASSESSMENT	
121	
SUMMARY.....	121
5.1 BACKGROUND	122
5.2 THEORETICAL FOUNDATIONS.....	124
5.3 METHODS	127
5.4 RESULTS – THE ADVANCE VALUE TREE.....	132
5.5 DISCUSSION - THE ADVANCE VALUE FRAMEWORK.....	149
5.6 CONCLUSION.....	155
CHAPTER 6 – PAPER 4	159
APPLICATION OF THE MCDA VALUE FRAMEWORK: A SIMULATION EXERCISE ON METASTATIC COLORECTAL CANCER WITH MULTIPLE STAKEHOLDERS IN THE ENGLISH SETTINGS	
159	
SUMMARY.....	159
6.1 BACKGROUND	160
6.2 METHODS	161
6.3. RESULTS.....	175
6.4 DISCUSSION	188
6.5 CONCLUSION.....	191
CHAPTER 7 – PAPER 5	193
APPLICATION OF THE MCDA VALUE FRAMEWORK: A CASE STUDY ON METASTATIC PROSTATE CANCER WITH THE SWEDISH DENTAL AND PHARMACEUTICAL BENEFITS AGENCY TLV	
193	
SUMMARY.....	193
7.1 BACKGROUND	194
7.2 METHODS	196
7.3 RESULTS.....	209
7.4 DISCUSSION	216
7.5 CONCLUSION.....	218
CHAPTER 8 – CONCLUSIONS, POLICY IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH DIRECTIONS	219

8.1 KEY CONCLUSIONS	219
8.2 POLICY IMPLICATIONS	227
8.3 LIMITATIONS	233
8.4 FUTURE RESEARCH DIRECTIONS	241
APPENDIX TO CHAPTER 2.....	278
APPENDIX TO CHAPTER 5.....	279
APPENDIX TO CHAPTER 6.....	280
APPENDIX TO CHAPTER 7.....	304
APPENDIX TO PUBLISHED ARTICLES	319

List of Chapters

Chapter 1 – Introduction to Health Technology Assessment, Thesis Research Questions and Overview

Chapter 2 – Current Methods of Value Assessment of Medical Technologies

Chapter 3 – Conceptual Framework: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment

Chapter 4 – Evaluation Criteria and Decision-Making Practices in Health Technology Assessment: a Systematic Review and Expert Consultation across 8 EU Countries

Chapter 5 – Methodological Framework: a Multiple Criteria Decision Analysis Value Measurement Model and Techniques for the Evaluation of New Medicines in Health Technology Assessment

Chapter 6 – Application of the MCDA Value Framework: a Simulation Exercise on Metastatic Colorectal Cancer with Multiple Stakeholders in the English Settings

Chapter 7 – Application of the MCDA Value Framework: a Case Study on Metastatic Prostate Cancer with the Swedish Dental and Pharmaceutical Benefits Agency TLV

Chapter 8 – Conclusions, Policy Implications, Limitations and Future Research Directions

List of Figures

Figure 3.1: Value function for scoring the performance of alternative options

Figure 3.2: Multiple criteria decision analysis (MCDA) methodological process in the context of health technology assessment

Figure 3.3: Value tree hierarchies and data sources using a multiple criteria decision analysis (MCDA) process for value assessment

Figure 4.1: Main components and sub-components of the analytical framework applied

Figure 4.2: Flow chart of the systematic literature review process

Figure 5.1: The five-stage criteria selection process for structuring the value tree as part of a top-down ‘value-focused thinking’ approach.

Figure 5.2: The generic value tree for new medicines Advance Value Tree

Figure 6.1: Preliminary value tree for metastatic colorectal cancer (pre-workshop)

Figure 6.2: Final Value Tree for metastatic colorectal cancer (post-workshop)

Figure 6.3: Stacked bar plot of treatments’ overall weighted preference value (WPV) scores across all attributes

Figure 6.4: Criteria weights histogram

Figure 6.5: Cost benefit plot of overall weighted preference value (WPV) scores versus costs

Figure 7.1: Preliminary value tree for metastatic prostate cancer (pre-workshop)

Figure 7.2: Final value tree for metastatic prostate cancer (post-workshop)

Figure 7.3: Stacked bar plot of treatments’ overall weighted preference value (WPV) scores across all attributes

Figure 7.4: Criteria weights stacked bar

Figure 7.5: Cost benefit plot of overall weighted preference value (WPV) scores versus costs

Figure 8.1: Differences between cost effectiveness analysis (CEA) and multiple criteria decision analysis (MCDA) in decision-making

Figure 8.2: Proposed roadmap to MCDA application and implementation

List of Tables

Table 3.1: Different methodological approaches for selecting objectives and criteria

Table 4.1: Responsibilities and structure of national HTA agencies

Table 4.2: Evidence and evaluation criteria considered in HTAs

Table 4.3: HTA methods and techniques applied

Table 4.4: HTA outcomes and implementation

Table 5.1: Value dimensions considered as evaluation criteria among the study group of European countries and their intensity of use

Table 5.2: Comparison of the Advance Value Framework with other value frameworks

Table 6.1: Attributes definition and sources of evidence

Table 6.2: Pre-workshop attribute reference levels and basis of selection

Table 6.3: List of decision conference participants

Table 6.4: Changes in the definitions of the attribute reference levels, pre- and post-workshop

Table 6.5: Post-workshop attribute reference levels and basis of selection

Table 6.6: Options performance across the criteria attributes

Table 6.7: Overall weighted preference value (WPV) scores, partial preference value scores, relative weights, costs and cost per unit of value

Table 7.1: Attributes definition and sources of evidence

Table 7.2: Pre-workshop attribute reference levels and basis of selection

Table 7.3: Options performance across the criteria attributes

Table 7.4: Overall weighted preference value (WPV) scores, partial preference value scores, relative weights, costs and cost per unit of value

List of Boxes

Box 1: Definitions of decision analysis terminology

List of Abbreviations

AEs	Adverse Events
AIFA	Agenzia Italiana del Farmaco (Italy)
AMNOG	Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Germany)
AOTMiT	Agency for Health Technology Assessment and Tariff System (Poland)
ASMR	Amélioration du Service Médical Rendu (HAS/ France)
ATC	Anatomical Therapeutic Chemical
BIA	Budget Impact Analysis

BNF	British National Formulary
BoD	Burden of Disease
BSC	Best Supportive Care
CBA	Cost Benefit Analysis
CEA	Cost Effectiveness Analysis
CED	Coverage with Evidence Development
CEESP	Commission for Economic Evaluation and Public Health (France)
CEPS	Economic Committee for Health Products (France)
CPR	Pricing and Reimbursement Committee (Italy)
CUA	Cost Utility Analysis
DA	Decision Analysis
DH	Department of Health
EBM	Evidence Based Medicine
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal Growth Factor Receptor
EMA	European Medicines Agency
EOL	End of Life
EPAR	European Public Assessment Report
ERG	Evidence Review Group (NICE/ England)
EU	European Union
EUnetHTA	European network for Health Technology Assessment
FOLFIRI	Folinic acid, (5-)Fluorouracil and Irinotecan
FOLFOX	Folinic acid, fluorouracil and oxaliplatin
FP7	Seventh Research Framework Programme
G-BA	Federal Joint Committee (Germany)
GDP	Gross Domestic Product

HAS	Haute Autorité de Santé (France)
HRQOL	Health Related Quality of Life
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
ICP	Interministerial Committee for Pricing (Spain)
ICVR	Incremental Cost Value Ratio
INN	Innovation Level
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Germany)
ISCIII	Instituto de Salud Carlos III (Spain)
ISP	Intérêt de Santé Publique (France)
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LYG	Life-years gained
MACBETH	Measuring Attractiveness by a Categorical Based Evaluation Technique
MAUT	Multi-Attribute Utility Theory
MAVT	Multi-Attribute Value Theory
MCDA	Multiple Criteria Decision Analysis
mCRC	Metastatic Colorectal Cancer
mCRPC	Metastatic Castrate-Resistant Prostate Cancer
MEA	Managed Entry Agreement
MoH	Ministry of Health
NICE	National Institute for Health and Care Excellence (England)
NHS	National Health Service
NMTs	New Medical Technologies
ORR	Objective Response Rate
OS	Overall Survival
PFS	Progression Free Survival

QALY	Quality Adjusted Life Year
R&D	Research and Development
RCT	Randomised Control Trial
RedETS	Red de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (Spanish HTA network)
RoA	Route of Administration
SAF	Safety Profile
ScVJs	Scientific Value Judgments
SESCS	Servicio de Evaluación y Planificación Canarias (Spain)
SMR	Service Médical Rendu (HAS/ France)
SOC	Socioeconomic Impact
SoVJ	Social Value Judgement
SPC	Summary of Product Characteristics
SMAA	Stochastic Multi-Criteria Acceptability Analysis
SWF	Social Welfare Function
TA	Technology Appraisal
THE	Therapeutic Impact
TLV	Tandvårds- och Läkemedelsförmånsverket (Sweden)
VA	Value Assessment
VBA	Value Based Assessment
VBP	Value Based Pricing
WHO	World Health Organisation
WPV	Weighted Preference Value
WTP	Willingness to Pay
ZIN	Zorginstituut Nederland (The Netherlands)

Chapter 1 – Introduction to Health Technology Assessment, Thesis Research Questions and Overview

1.1 Introduction and the case for improving the way we measure the value of health care interventions

One of the foremost challenges health care systems are facing is the scarcity of resources in combination with rising demand for services, putting their sustainability in danger. As a result, decisions relating to the allocation of health care resources has been inevitable, either between different competing services and interventions (i.e. priority setting) or across different patients (i.e. rationing). However, the methodological approach of allocating resources in an efficient and fair way that gives legitimacy to the decision outcomes has been far from obvious⁴⁻⁹. This is in large part due to i) the complexity of the decisions, as a variety

of different factors and objectives need to be balanced through the involvement of a range of stakeholders, ii) the importance of the decision outcomes, as they have a dramatic impact on human health, and iii) the ethical and social responsibilities behind the provision of health care which traditionally has been perceived as a government duty, given that health on its own is regarded as a public good or even a human right.

1.1.1 Resource Allocation Methods in the National Health Service

The need for priority setting in health care had not always been realised; its acknowledgement came following a lengthy and incrementally evolving process characterised by the interplay of scientific advancements, culture changes and politics.

However, the process of priority setting in the last 20 years has remained unchanged. The example of the British National Health Service (NHS) provides an insightful case study. Nowadays, priority setting takes place across all levels of the organisational hierarchy of the health care system: the central government decides the overall budget of the NHS, commissioners and providers determine their purchases among alternative services and interventions, and clinicians allocate their time and resources ¹⁰. A similar decision context was evident in the 1990s. As Klein described in 1993, “micro-decisions about priority setting are constrained by macro-decisions about resource allocation taken at superior levels in the organisational hierarchy” ¹¹(p. 309), according to an almost identical landscape where cabinet decided on the NHS budget, Department of Health decided the priority targets, and purchasers decided on services ¹¹.

As a result of the multi-level context of priority setting, rationing can be implemented in various ways: rationing by deterrence, when obstacles to patient access are imbedded; rationing by deflection, when the responsibility of service provision is passed on to another agency; rationing by dilution, where the quality of service declines; rationing by denial, when a particular treatment is refused to get funded; and rationing by selection, when a treatment is only allowed for a particular population sub-group ^{10,12}. It seems that rationing by deterrence or delay was the first of these models to emerge, with Roy Parker already describing this approach in the mid-70s ¹³. All these rationing instruments were already applied more than two decades ago, possibly with the exception of the relatively harsher approaches of rationing by denial and selection which became more abundant due to increased fiscal pressures later on ¹¹.

The decision-making tool used for budget distribution has not changed either. Quality Adjusted Life Year (QALY), which is calculated by multiplying the value of each health state by the respective length of time of each state, provides in a single common summary measure the total health improvement, therefore providing a simple tool for resolving complex choices and turning it into the most widely used measure of health status for the assessment of health benefits ^{14,15}. However since then, it had been argued that the Quality Adjusted Life Year (QALY) is flawed as a way of priority setting in health care ¹⁶ and that it does not originate directly from individual's utility function therefore only partly reflecting the individual's true preferences ¹⁷, while it had been acknowledged that it is "beset by methodological problems about the valuation of different states of health, by lack of data about incomes, and by the problem of patient heterogeneity" ¹¹ (p. 310). Although it was more or less agreed that priority should ideally take place based on demonstration of effectiveness or need, both of which were problematic concepts however (at least back in that time), the former due to its uncertainty and the latter due to its ambiguity. And Klein therefore continued with the realisation that the process was not "rational" (probably in a logical or reasonable sense) but instead priorities were "emerging from pluralistic bargaining between different lobbies, modified by shifting political judgements made in the light of changing pressure" ¹¹ (p. 310). In doing so though he argued that given the multiplicity of conflicting values (i.e. objectives) brought along such discussions, the idea of "a machine grinding out priorities" and making decisions for us would be "absurd", and that the limitation of pluralistic bargaining was that it was not pluralistic enough but instead dominated by some (clinical) voices. As a consequence he finally argued that in order for the process to become more rational the technical characteristics of the decision-making process should be improved, according to an "open dialogue, [...], in which arguments can be tested against evidence and the conflicts between different values or preferences can be explored" ¹¹(p. 310), a rationality concept that he acknowledged goes back to Aristotle's work of finding "good reasons" to justify decisions ¹⁸.

Consensus has now been reached that emphasis should be placed on the process of decision-making in order to assess the efficiency and fairness of decision outcomes. Daniel's and Sabin's Accountability for Reasonableness ethical framework has been cited by most for this reason ¹⁹, which states that for decisions to be fair and legitimate their processes should meet a number of conditions: they should be publicly available, based on relevant reasoning, and revisable in light of new evidence, all three conditions of which could be assured through enforcement mechanisms (i.e. regulation) ²⁰. Given that allocation of

resources is also a political process, the application of such an ethical framework is very much needed.

1.1.2 Pharmaceuticals and Need for Value Assessment

Pharmaceutical spending is the third largest component of health care expenditure after hospital and ambulatory spending, accounting for between 10% and 20% of total expenditure on health in most OECD countries (the OECD average being 16.5% in 2010, or 1.5% of GDP)²¹. Faster uptake of new and more expensive drugs has caused an escalating trend in pharmaceutical spending per capita in many developed economies²².

Although new, branded medicines account for only about 25% of all drugs dispensed by volume, they account for 70% to 85% of pharmaceutical spending across most OECD countries²³; price growth is mostly attributable to innovative medicines in a relative small number of therapeutic categories such as cancer, infection and central nervous system products²².

Most expensive new medicines are associated with uncertainty in clinical benefit, alongside high cost. The associated clinical uncertainties usually relate to two aspects: first, safety due to potential unknown adverse drug events and, second, real effectiveness in daily clinical practice. Both of these uncertainty aspects could potentially be attributed to the design of the clinical research studies most of which fall short of a number of necessary features for their findings to be useful²⁴, such as their information gain (i.e. is the study large and long enough to be sufficiently informative?), pragmatism (i.e. does the study reflect real life?), and patient centeredness (i.e. does the study reflect patient priorities?). Failure to compare all relevant treatment options and the relatively short-term or ‘truncated’ time horizons of the studies act as additional limitations²⁵. As a result, the clinical uncertainty of new medicines in tandem with their high cost in prices have led to the assessment and appraisal of their benefits and costs by the payers in order to estimate their value and determine whether they should be covered, and possibly at what price²⁶⁻²⁸.

In contrast to commodity markets, such as those for generic medicines, the price mechanism in more complex markets such as the one of innovative medicines has a more diversified role: besides reflecting the actual value of a product to consumers in the present time, i.e. static efficiency, it also needs to encourage future investments and the potential societal gains emerging from them in the future, i.e. dynamic efficiency²⁸. In this context, the emergence of value based pricing (VBP), highlights the need to set drug prices after

assessing their value to a variety of stakeholders, such as patients, health care systems and even society at large ^{27,29,30}. One of the main purposes of VBP is to essentially improve patient access to effective and innovative drugs while ensuring their prices are reflective of their value ³¹.

Central to VBP is the application of Health Technology Assessment (HTA) processes that evaluate the clinical benefits, and usually, costs of new health care interventions (including drugs, diagnostics and medical devices), compared to existing available alternatives in order to inform pricing and reimbursement decisions. Despite the extensive uptake and use of HTA practices and processes in recent years, there are several conceptual and methodological issues inherent in the use of the current value assessment (VA) approaches that limit their success.

The implication of ever increasing pharmaceutical expenditure in combination with the finite amount of resources available for health care is that health care decision-makers such as payers and commissioners of care, need to allocate resources efficiently, while ensuring patient access to effective new medicines. This observation is one of the central drivers of the thesis. It is suggested that the problematic nature and the technical difficulties associated with assessing the value of new medical technologies (NMTs) in practice, including their link to decision-making, are the key obstacles to overcome in order to address the above issues.

1.1.3 Primer on Economic Value Theories

A brief background of various value propositions in a historical context is useful to understand the evolution of economic value theory which could be generally classified into pre-classical, classical and neo-classical thinking. In medieval times, “value” was generally perceived by theorists to depend on utility and scarcity, with a subjective approach to value being generally adopted ^{32,33}. It was in the 17th century, when the pre-classical thinking on value started to emerge, that cost of production started to be considered as part of the value function ^{32,34}. According to William Petty, the market price of a commodity would fluctuate around its “natural value” that would depend on land and labour (the factors of production), therefore proposing a more natural or objective theory of value ^{32,35}. Petty’s theory was eventually reduced to a labour explanation expressed in labour cost, therefore representing a labour cost theory of value. According to a similar - but not identical - logic, Richard Cantillon proposed to express cost of production in units of land, therefore adopting a land theory of value ^{32,36,37}. Ferdinando Galiani was among the pre-classical economists that

continued to support a subjective, utility-based, approach to value evident from the medieval period, the utility theory of value ³²⁻³⁴. John Law first proposed the use of a two dimensional approach considering both demand and supply factors to determine the value of a good in society, based on which he tried to explain the water-diamond paradox of value as part of which the high value of water, essential for life, was priced at lower market price to the value of diamond, inessential to human life ³⁸. By distinguishing between “value in exchange” and “value in use”, he regarded that diamonds’ “high value in exchange” is due to their relative scarcity, his reasoning essentially fitting the marginal utility approach that ultimately resolved the paradox in the 19th century turning his work into a milestone ³⁹.

The seminal work of Adam Smith signalled the beginning of the classical thinking, pushing back towards Petty’s labour theory of value, compared to other pre-classical economists ⁴⁰. After realising that a labour theory would not hold in an advanced economy, he proposed a cost of production value theory consisting of elements from labour, land and capital theories. David Ricardo then adopted Smith’s rejected labour hypothesis proposing that in free markets the value of (freely reproducible) commodities is derived from their scarcity and the quantity of labour needed for their production ⁴¹. Karl Marx subsequently built on Ricardo’s labour theory of value, which supported that commodities can be viewed as functions of labour time ⁴². Although John Stuart Mill allegedly also continued to work Ricardo’s labour theory, he recognised that a commodity’s market value is related to both demand and supply effects, placing his value approach closer to the neo-classical school of thought ⁴³.

The latest era of neo-classical thinking was marked in the 19th century by the work of William Stanley Jevons, Carl Menger and Leon Walras who almost simultaneously developed the idea of marginal utility, trying to find a cause and effect relationship between value and utility that excluded the cost of production ⁴⁴⁻⁴⁶. Their work solved the water-diamond paradox of value, essentially supporting the notion that value entirely depends on utility as Jevons had said, and more precisely on marginal utility. In contrast to Jevons, Menger did not believe that goods provide units of utility, or “utils”, but instead that goods value is derived because they serve various uses whose importance differs and which satisfy peoples wants ⁴⁵. It was Leon Walras and Alfred Marshall that later combined together both utility and cost of production (i.e. demand and supply) theories to produce a complete two-dimensional value theory ^{47,48}. Walras’ biggest contribution was the General Equilibrium Theory, attempting to demonstrate how a whole economy works together and reaches equilibrium as part of a complex economic system integrating both the demand and supply

side factors ⁴⁶. Marshall proposed a Partial Equilibrium concept and emphasised that the price of a good is determined both by supply and demand factors, and also introduced the price elasticity of demand concept, along with the consumer and producer surplus concepts ⁴⁹. Marshall's concept of externalities was then developed further by Arthur Cecil Pigou, arguing that the existence of externalities is sufficient justification for government intervention, advocating the use of taxes to discourage activities that lead to negative externalities, and the use of subsidies to encourage activities that created positive externalities ^{50,51}. The above economic value theories and in particular utility theory were then extended by and used in the field of decision analysis which emerged in the 1960s.

1.1.4 Utility Theory and Value Theory in Decision Analysis

Decision analysis was originally defined by Howard as “a logical procedure for the balancing of the factors that influence a decision” incorporating “uncertainties, values, and preferences in a basic structure that models the decision” ⁵². Preferences reflect how desirable or choice-worthy an option or characteristic might be over another alternative option or characteristic, usually interpreted in a comparative way. By definition, decision-making is inherently subjective as it depends on individual utility and preferences which differ among individuals. The logic behind DA was then described by Raiffa as “divide and conquer”: “decompose a complex problem into simpler problems, get one's thinking straight on these simpler problems, paste these analyses together with logical glue, and come out with a program of action for the complex problem” ⁵³.

The main outcome of decision analysis theoretical axioms is that first, the utility of an alternative is the indication of its desirability, and second, that an alternative A with higher utility $U(A)$ should be preferred to an alternative B with lower utility $U(B)$, in other words expressing the rationality rule of utility maximisation, i.e.

$$A \succ B \Leftrightarrow U(A) > U(B) \quad (1)$$

Decision analysis seems to be particularly useful for coping with the complexities arising from uncertainty and multiple conflicting objectives ⁵³. Uncertainties can be traded-off against some value aspects of the outcomes, formally through the incorporation of probabilities. Similar trade-offs can be made among different objectives and their associated

values. These trade-offs are judgements, depending on the decision maker's assessment of the relative desirability of the available options across their dimensions in tandem with the relative importance of these dimensions. However, given that trade-offs are personal, there are no universal rules for making them, being subjective in nature ⁵⁴. Therefore, the notion of rationality is used, with the goal of making rational inferences and decisions. Possibly the most prominent of such criteria rules would be the maximisation of (expected) utility or value ⁵⁵.

In decision analysis there is a clear conceptual difference between value functions, which assess the marginal benefit of an option, and utility functions, which incorporate preferences, therefore assessing both marginal benefit and risk attitude. The former is employed for riskless choices and the latter for choices under uncertainty. Expectation mainly relates to the concept of probability theory, with *expected* utility or value being a weighted average of the pay-off for an outcome (utility or value) and its respective probability of actually occurring ⁵⁴. Expected utility and expected value are, therefore, calculated in a similar way but the pay-offs in the former case correspond to subjective utilities rather than objective quantities, with their formal terminology being *subjectively expected utilities* (SEUs): both the utilities and the probabilities incorporated are numbers, but are subjective in nature giving rise to *numerical subjectivity*, the notion of subjective judgments expressed as numbers ⁵⁶. In the context of HTA, the evaluation of new medical technologies predominantly relates to the evidence-based assessment of their value by measuring their marginal benefits and therefore we choose to use the *value* term rather than *utility*, which reduces the heterogeneity of the value judgements based on expected utility.

The expected value (EV) of an event could then be written as

$$EV = p_1v_1 + p_1v_2 + p_3v_3 + \cdots + p_nv_n \quad (2)$$

Or alternatively as

$$EV = \sum_{i=1}^n p_i v_i \quad (3)$$

Where p_i is the probability that event i will take place and v_i is the value or pay-off associated with the event.

Given that the evaluation of health care interventions as part of HTA predominantly relates to the assessment of their value using existing evidence and not expected evidence, for simplicity reasons we choose to detach expectation and the probability concept while assuming the existence of evidence for the measurement of marginal benefit. However, it could also be argued that in various cases absence of (satisfactory or adequate) evidence essentially introduces an expectation variable given the attached probabilities of the respective outcomes to take place (e.g. clinical outcome), especially as part of early-HTA settings where a new medicine might still be under clinical development.

Overall, based on *value theory* alternative a is preferred over alternative b , if and only if $V(a) > V(b)$, and is judged to be indifferent if and only if $V(a) = V(b)$, where V is a real number reflecting the value associated with the performance of the alternatives based on which preference orderings are produced^{2 54}. In *multi-attribute value theory* (MAVT), which is essentially the extension of value theory for the evaluation of problems with multiple value dimensions, partial value functions for each criterion, for example $v_i(a)$, are first constructed and then aggregated together. Once again, the partial value function must satisfy the previous definition of a preference function according to which an alternative a is preferred to an alternative b for criterion i , if and only if $v_i(a) > v_i(b)$, with a similar indifference working hypothesis existing only if $v_i(a) = v_i(b)$. Importantly, the working assumptions required for the formation of the partial value functions are interrelating with the aggregation type used².

Typically, an additive (linear) aggregation approach is adopted, taking the following form:

$$V(a) = \sum_{i=1}^m w_i v_i(a) \quad (4)$$

Where m is the number of criteria, and $w_i v_i(a)$ the weighted partial value function of criterion i for option a . Normally, partial value functions are standardised using lower and higher reference levels at 0 and 100 corresponding to “worst” and “best” or “least preferred” and “most preferred” outcomes. The “weight” w_i represents the relative importance of criterion

² Assuming preferences are complete (i.e. for any pair of options, either one is strictly preferred to the other or they are indifferent) and transitive (i.e. if an alternative A is preferred to an alternative B, and if B is preferred to an alternative C, then A is preferred to C)

i , which can be conceived as a parameter reflecting the gain associated with the replacement of the worst outcome by the best outcome for this particular criterion ², which should be elicited using weighting techniques that assess their value trade-offs. Typically, the quantitative swing weighting technique is used asking for judgments of relative value between ‘swings’ (i.e. changes) from standard lower level x_* to higher reference level x^* on each x -th attribute), with each swing (i.e. a relative change from a lower to a higher attribute level) being valued between 0 and 100 and the most valuable swing anchored as 100 ⁵⁴. Normalised weights are then calculated, as a proportion of each swing weight, so the normalised weights summed to 100%. Overall, specification of the additive model requires the use of explicit reference values of 0 and 100 and positive weights summing one (or 100%). The additive model is assumed as a working hypothesis and therefore the model construction is done to respect these properties underlying the additive value model.

In order to enable the use of simple aggregation rules (e.g. additive value models, where scores and weights of the different individual criteria are multiplied and then added altogether in a weighted average manner), preference independence between the different criteria needs to be upheld ². Preference independence is a key property; it implies that an option’s value score on a criterion can be elicited independently of the knowledge of the option’s performance on the remaining criteria. It should be noted that preference dependence (cardinal and ordinal) is not the same as environmental dependence (e.g. statistical correlation); two different criteria attributes could be statistically correlated (i.e. environmentally dependent) but at the same time preference independent and vice-versa. If this requirement is not observed, an additive aggregation function should not be employed unless the criteria are restructured to combine non-preference-independent criteria into a single criterion.

Usually, simpler aggregation rules such as the additive model are preferred over more complex rules such as multiplicative models for merging scores and weights. This is mainly because of their simplicity and comprehensible nature which is easily explained and understood by any decision-makers ², but also the considerably smaller errors associated with its use compared to the errors coming from the incorrect aggregation of preferences through more advanced models as evident on empirical evidence from simulation studies ⁵⁷. Eventually, if the above requirements cannot be satisfied, then more complex aggregation rules would have to be applied to combine scores with weights. In turn, other aggregation and disaggregation approaches are available in the DA literature but this thesis is not covering them⁵⁸.

1.1.5 Problematic

Due to the different levels at which resource allocation needs to take place, and the fact that priority setting is a process focusing on the general population in contrast to the process of rationing which focuses on individual patients, a mix of slightly opposing principles act as objectives for resource allocation ¹⁰. On one hand there is the purely utilitarian principle of maximising the health impact on the whole population ⁵⁹, and on the other hand there is a set of, usually secondary, ethical objectives relating to the distribution of health that mainly aim to prioritise interventions which target the more vulnerable ⁶⁰, such as the poor ⁶¹, seriously sick ⁶², women and children ⁶³.

These objectives should in theory be operationalised through the application of a plurality of criteria, most of which are characterising either the intervention under consideration or the condition it is indicated for. The intervention is usually assessed through the notion of benefit risk-ratio, essentially an evaluation metric reflecting whether the benefits of implementing the intervention outweigh the risks, by accounting both for the impact on health and the impact on resources needed. The condition is assessed through its burden or severity, which is usually approximated through its seriousness (morbidity and mortality related) and the availability of treatments.

As Baltussen and Niessen have described, a number of priority setting approaches have already been developed over the last 20 years, but all of these tend to concentrate on single value dimensions ⁶⁴. These include the “evidence based medicine” approach for prioritising interventions according to their established effectiveness, “cost-effectiveness analysis” (and other forms of full economic evaluation) for prioritising according to efficiency by accounting both for effectiveness and costs, “equity analyses” for prioritising according to distributional impact, and “burden of disease analysis” for prioritising diseases according to their burden (through morbidity and mortality).

Most of the current values assessment approaches adopted as part of HTA mainly consider (comparative) clinical efficacy in combination with or without clinical cost-effectiveness techniques, while increasingly incorporating real world data after the drug has entered the market, thus essentially reflecting comparative effectiveness and efficiency ²⁸. However, there is considerable arbitrariness in the selection of evaluation criteria used to interpret evidence and determine value, including which metrics to use for measuring

efficacy and effectiveness, what type of costs to consider, and very importantly how to account for other key dimensions of benefit.

The value of NMTs is multi-dimensional and not only limited to clinical benefit and cost but spans the disease severity and target population size, the nature of the intervention and whether for example it is curative or preventive, economic impact and budgetary constraints, together with other factors such as evidence quality ⁶⁴⁻⁶⁶. Thus, the methodological framework of these value assessment approaches is inadequate and at best partial mainly because the evaluation criteria used to assess evidence and determine value are incomplete. Many important value considerations falling under the burden of disease the treatment addresses, the treatment's innovation level and its overall socioeconomic implications, are not adequately reflected in the evaluation process. Typically, as part of traditional economic evaluation techniques these value dimensions are not considered, and if they are, this might be done in an implicit and non-methodical manner, by considering a value concern only informally on an *ad-hoc* basis through committee deliberation and possibly as part of non-transparent negotiations with the manufacturer. For example, such an *ad hoc* approach would characterise the evaluation of expensive life-extending end-of-life (EOL) treatments in the English context ⁶⁷, which has generated important questions about the credibility and consistency of the whole process ⁶⁸. In addition, there are technical issues in achieving consensus on value, for example, what should the importance of each criterion be and how to account for uncertainty and incomplete evidence. Essentially, an explicit definition of value that relies on a comprehensive set of parameters is missing ⁶⁹.

Due to the complexity of these multi-criteria problems, decision-makers tend to adopt intuitive or heuristic approaches for simplification purposes, but as a consequence important information might get lost or not utilised leading to choices based on an *ad hoc* priority setting process ⁶⁴. Under these conditions, requiring multiple trade-offs across a range of societal values, decision-makers seem not to be well equipped to make well-informed, "rational" decisions ^{70,71}. Therefore the need has arisen for a transparent approach that can help decision-makers to maximise societal welfare through better resource allocation decisions across the field of health care. Decision-makers have shown interest in incorporating additional dimensions of value through the use of multi-criteria methods ⁷². These include the European Medicines Agency (EMA) for benefit risk assessment ⁷³, the Institute of Medicine in the US for prioritising vaccines ⁷⁴, the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany for distinguishing between multiple clinical endpoints ⁷⁵, as well as the National Institute for Health and Care Excellence (NICE) in

England for the assessment of ultra-rare therapies ⁷⁶. As a result, it would be expected that decision-makers and other stakeholders would benefit from clear and comprehensive ways that allow them to assess all critical value dimensions of new health care interventions, in order to make better decisions about priority setting. Not having such methods creates a conceptual, methodological and policy gap.

A more explicit definition of value could incorporate a wider set of parameters than is currently the case, in combination with a more comprehensive methodological approach of assessment that includes views from all relevant stakeholders. This could then lead to the development of a supporting tool for decision-making, being flexible enough to enable decision-makers exercise their judgment, help them weigh the trade-offs and elicit their preferences when pursuing multiple objectives, while contributing to the debate on more efficient resource allocation.

1.2 Research Questions and Thesis Overview

The thesis' working hypothesis is that a value assessment system for health care technologies characterised by comprehensiveness, consistency and transparency could potentially enable better prioritisation decisions for problems with complex information. Eventually, this could lead to a better understanding of the decisions' rationale, therefore enhancing their acceptability.

1.2.1 Research Questions

The main research question of this thesis is the following: 'How to develop, if possible, a methodological framework that assesses the overall value of new medical technologies in a universal way by capturing aspects of value that are not explicitly addressed by other available methodologies? Could Multiple Criteria Decision Analysis (MCDA) provide a useful alternative for value assessment and what techniques or options are available for the context of HTA? Given that the above outlined research hypothesis is based on the application of MCDA methodology, the need to address a number of secondary research questions arises, which forms the basis for the empirical chapters of the thesis and proposed research papers, notably:

- 1) How to design a robust methodological process subscribing to MCDA principles for assessing the value of NMTs in HTA?

- 2) What are the current practices, processes and policies of value assessment for new medicines across a number of jurisdictions in Europe that employ explicit evaluation frameworks as part of their HTA processes and what are the parameters of value acting as evaluation criteria?
- 3) How to develop a sound MCDA methodological framework based on decision theory for the evaluation of new medicines that incorporates and models a multitude of value parameters for use in HTA?
- 4) How to adapt and apply an MCDA value framework in practice to rank a set of medical technologies according to their value by eliciting the preferences of stakeholders, possibly linking their elicited value scores with hypothetical coverage decisions?

1.2.2 Thesis Overview

MCDA methods have been proposed for use in the field of public services¹ such as transport^{77,78} and health care^{64,79,80}, including for priority setting of health programmes or interventions⁸¹⁻⁸⁴ and the regulatory approval of pharmaceuticals⁸⁵⁻⁸⁷, but has not been successfully implemented for the purpose of measuring the value of NMTs as part of HTA. However, MCDA could be used for value assessment at HTA level because of its potential comprehensiveness in accommodating a multiplicity of criteria with distinct weightings to assess value, its encompassing nature in terms of extensive stakeholder engagement, and transparency. Only very recently have others advocated the use of MCDA for HTA, recognising its potential value in the appraisal process and arguing that it could be used as an aid to decision-making, given that methodological considerations are addressed^{68,88-91}. Still, various aspects of such a methodology remain to be advanced and become aligned with current HTA systems needs for its use to be embraced by decision-makers, as for example what evaluation criteria and techniques to use (and how to choose from), and how to link results with policy making decisions.

Following Chapter 1, the thesis introduces current methods of value assessment in Chapter 2, including their perceived limitations. Five research papers are undertaken to address the overarching thesis question as part of Chapters 3-7 and which relate to each of the subsidiary research questions respectively, forming the main empirical body of the thesis. Research question 1 is addressed as part of Chapter 3 (Paper 1), research question 2 is

addressed as part of Chapter 4 (Paper 2), research question 3 is addressed as part of Chapter 5 (Paper 3), with research question 4 addressed as part of Chapters 6 and 7 (Papers 4 and 5).

Paper 1

Paper 1 builds a generic methodological process to act as the conceptual basis for the development of a robust MCDA value framework in the context of HTA. It provides an introductory classification of different MCDA schools of thought before suggesting a specific methodological process for applying such methods in HTA. Overall, it aims to bring together theoretical foundations of MCDA with HTA application, thus possibly acting as a manual of good practice, while also suggesting different means of linking its use with policy-making.

The generic process of MCDA includes a number of common stages that would have to be adapted to the context of health care and HTA ^{1,92}:

- a. Establish the decision context. Define the aim of the analysis, who are the decision-makers and any other key players;
- b. Establish objectives that indicate the overall purpose to be achieved and identify criteria of favourable and unfavourable effects;
- c. Identify the options to be evaluated against the criteria;
- d. Assign ‘scores’ to the options by assessing the performance of each option against the criteria and their value.
- e. Assign ‘weights’ to the criteria by assigning relative weights to each of the criteria to reflect their relative importance to the decision.
- f. ‘Aggregate’ scores and weights for each of the option by combining them together to derive overall value scores.
- g. Analyse the results.
- h. Conduct a sensitivity analysis to validate the results by testing the influence of changes in the scores or weights.

This process seems to be highly relevant to HTA and thus its application to health care settings; a decision perspective is initially decided, candidate technologies are assessed against pre-defined criteria, and results are then examined. However, a compelling adaptation has to be made to reflect the special case of the HTA context alongside the associated conceptual/methodological issues.

Overall, in order for the above tasks to be completed effort is needed in studying and learning the theoretical foundations of advanced decision sciences, understanding the characteristics of different MCDA methods, and reviewing past MCDA applications in various sectors for the development of a new conceptual methodological process that is adapted for the needs of HTA and in alignment with decision theory principles.

Paper 2

Paper 2 systematically reviews and synthesises evidence on the value measurement methods and techniques applied across a number of jurisdictions in Europe that employ HTA processes, focusing on the different evaluation criteria employed. The aim is to critically review, through literature and expert consultation, the value parameters considered as part of current methods applied for assessing and appraising the value of new medicines, with the view to incorporate these dimensions as value domains into the next paper, thus forming a model that feeds the value framework.

The application of MCDA in other contexts, such as transport, is relatively straightforward because the main aim of the analysis is usually conceptually easy to understand, and thus the appropriate criteria are easily identifiable. For example, when encountered with the objective of deciding on the location for the construction of a new railway, it would be expected to end up with criteria reflecting the following concerns: urban and environmental impact, integration with the rest of the transport network, complexity of construction and costs ⁹³. In contrast, in the field of HTA, the concept of “value” is much more complicated and a complete set of criteria needs to be identified and adapted to describe the value of a new medical technology in a specific context or disease area. Consequently, a significant contribution is needed in order to establish a definition of value, by considering all possible and relevant value dimensions of NMTs and selecting the right attributes to describe those dimensions.

The databases of PubMed and Social Science Citation Index (through the Web of Science portal) are searched for peer review literature published in English³ using the following keywords: ‘value assessment AND pharmaceuticals/methodologies criteria’ and ‘health technology assessment AND pharmaceuticals/ methodologies criteria’. A review protocol for identifying the relevance of articles is used for inclusion and exclusion of the various studies. Additional input is included from the policy and grey literature. The

³ Study countries also publish reviews and recommendations in English.

literature findings are validated and further insights are obtained from consultations with policy experts and using semi-structured interviews.

Understanding the methodologies that exist and the manner in which the criteria for value assessment are implemented in different HTA settings is required to identify the current limitations of existing value frameworks and the metrics of valuing NMTs from a comparative perspective; evidence of additional parameters that contribute to the perceived value of technologies, such as the burden of disease, the innovation level of the treatment, and wider socioeconomic implications, could inform the production of a robust methodological approach for more encompassing value assessment. The above critical review of the literature, in combination with expert communication, is used to identify, and subsequently feed into the development of a more comprehensive methodological framework for value assessment, by helping to identify and select a complete set of criteria that entirely describe the “value” of NMTs. In this context, the implementation of an MCDA framework can be a logical way forward.

In sum, achievement of the above tasks involve collecting both secondary and primary data, conducting a systematic literature review by studying specific literature of different countries and HTA agency guidelines, combining evidence together, seeking expert feedback, conducting semi-structured interviews, validating evidence and analysing results.

Paper 3

Paper 3 outlines the development of the value framework including the design of a value assessment model taking the form of a value tree. It commences with the theoretical axioms of Decision Analysis (DA), introducing the logic behind DA and the use of MCDA as a decision-making tool. It then links the findings of the systematic review and expert consultation with the new conceptual methodological process (Papers 2 and 1 respectively) for the design of a value tree, informing the selection of criteria needed to capture the value of new medicines. Finally, it outlines different types of MCDA methods, before suggesting a combination of techniques to operationalise the value tree so that it can be used for evaluating new drugs in HTA.

The MCDA value framework is generic for NMTs but focuses on the application of new medicines, illustrating the adaptation of the two initial phases of the MCDA methodological process from Paper 1, together with a recommendation for the use of specific MCDA techniques, notably:

1. Establish the decision context. Define the aims of MDCA, and who the decision-makers and other key players are;

For the purpose of value-based assessment of NMTs, the decision context could adopt a societal point of view that aims to maximise social welfare. Key stakeholders involved could be regulators, health care professionals, methodology experts, patients, and possibly manufacturers.

2. Identify the objectives or criteria that capture value relevant to the decision-making context;

Possibly the most fundamental step affecting the MCDA model and the results of the analysis is to decide on the evaluation criteria for which the options are scored against. These criteria, or grading parameters, represent factors that are important when assessing the “value” of a drug in a VBP context, and could be chosen through the critical review and expert consultation carried out in the previous paper of the thesis and the limitations identified therein. The identification of a complete set of evaluation criteria with the objective of exhaustively describing all potential dimensions of value has not been successfully undertaken yet and forms a key intellectual contribution of the thesis. In particular, drug–indication characteristics are grouped into clusters addressing a particular treatment characteristic. Criteria clusters (groups) are constructed according to decision analysis theory on the principles of completeness (non-excludability), non-redundancy (parsimony), operationability (usefulness), and mutual independence of preferences (preference scores on one criterion are not affected by preference scores of the rest criteria)¹. Five such key value domains, forming the relevant criteria clusters, are identified: (a) burden of disease, (b) therapeutic impact, (c) safety profile, (d) innovation level, and (e) socioeconomic impact. These areas form the higher or top-level criteria levels where the most important trade-offs take place and each of them in turn is made up of a number of more specific lower or sub-level criteria whose hierarchical representation takes the form of a value tree. A DA software is used (M-MACBETH) as a decision support system and tool to organise the clusters of evaluation criteria into a value tree which forms the basis of model, and also for scoring the options, weighting the criteria, analysing the results and conducting sensitivity analysis.

3. Provide an overview of MCDA modelling techniques, in regards to scoring, weighting and aggregating;

A recommendation of particular MCDA methods and precise combination of techniques for the stages of ‘scoring’, ‘weighting’, and ‘aggregation’ is provided, completing the development of a new value framework.

From a policy perspective, this paper aims to provide a new methodological framework for assessing the overall value of new medicines from a societal perspective, while providing an explicit and complete set of evaluation criteria.

To undertake and complete these tasks, effort is needed in understanding the use of advanced decision science methods and techniques, reviewing past MCDA applications, collecting secondary and primary data, conducting multiple rounds of literature reviews, seeking expert feedback, presenting the results and consulting with experts, combining evidence together and analysing results and linking decision science theory with HTA practice, so that a new methodological framework is developed that can serve the needs of decision-makers.

Papers 4 and 5

Paper 4 and Paper 5 constitute two separate case studies where the previously described and developed new MCDA framework is applied in practice to two different disease indications, namely (a) metastatic colorectal cancer (mCRC) and (b) metastatic castrate-resistant prostate cancer (mCRPC). Each of the papers involves an empirical testing of the methodological framework as well the adaptation of the generic value tree developed in Paper 3. Different drug treatments for each disease indication are used as the alternative options for the analysis, which are eventually assessed and compared with each other. Drugs are scored across a set of criteria, criteria are weighted according to their relative importance and overall value scores are produced, based on which the alternative treatments are ranked.

Cancer acts as a suitable example to describe the situation where difficult choices in resource allocation have to be made: new medicines for a life threatening condition and possibly with relatively incremental benefits across a range of value dimensions are coming in the market at much higher costs. Both case studies look at highly prevalent cancers for which there are available multiple expensive treatments, therefore forming the conditions of a problem that requires the assessment of value before an informed and rational choice is made. The first case study (Paper 4) relates to mCRC, eliciting preferences from a group of different experts and stakeholders in an attempt to mimic the composition of evaluation committees as in the case of the Technology Appraisal committees of the National Institute for Health and Care Excellence (NICE), the national HTA agency in England and Wales.

The second case study (Paper 5) relates to mCRPC, however preferences are elicited from the perspective of decision-makers, by working together with a group of evaluators from the national HTA agency of Sweden (TLV).

The methodology for both papers is based on the contents of Paper 3, outlining the development of the generic methodological framework for value assessment in the context of HTA. The value framework is adapted to the circumstances of each of the two disease conditions and the treatments available for each of them in order to be able to fully reflect their value, for example by incorporating specific evaluation criteria. Ultimately the aim for both papers is to test the newly developed methodological framework in practice.

Both primary and secondary data are used. Secondary literature sources (PubMed, Social Science Citation Index, and publicly available HTA appraisal reports) inform important value aspects of the drugs assessed which influence the adaptation of the general MCDA model developed in Paper 3 (e.g. selection of disease-specific clinical and surrogate endpoints). Primary data sources (i.e. expert opinion) are used for the validation of the model, and for the scoring and weighting steps of the process through a DA facilitated-workshop with stakeholder participation.

More specifically, for the purpose of scoring and weighting a facilitated workshop in the form of a decision conference is organised. In the first case study (Paper 4), key stakeholders participate including health care professionals (e.g. clinicians, pharmacologists, public health specialists), methodology experts (e.g. health economists, statisticians) and patient advocates (e.g. representatives from patient organisations), where a proportion of participants comes from existing or past NICE appraisal committees members that are responsible for the appraisal of cancer drugs. Approximately 7-15 experts should participate in the workshop, a number which has been shown to be ideal because it tends to preserve individuality while also allowing group processes; in other words, it is small enough to allow participants to reach an agreement but sufficiently large enough to represent all perspectives and interests ⁹⁴. On the other hand, for the purpose of the second case study (Paper 5), a smaller number of participants would be needed (e.g. 3-6 experts), given that all of them would come from the same stakeholder group being affiliated with the Swedish HTA agency.

An impartial facilitator (DA expert) coordinates the workshop and guides its process without contributing to it, intervening as appropriate to ensure the participating group remains task-oriented and achieves its objectives in the time available ⁹⁵. The workshop starts with an outline of the objectives and a presentation of the evidence available for the

different options relating to the assessment criteria. Initially the existing model is validated (as adjusted already through evidence from the literature), possibly involving on-the-spot amendments. Once the structure of the model is agreed upon, participants are invited to score the performance of the various options against the criteria and weight the relative importance of the criteria. Finally, results are presented. The DA software M-MACBETH is used for the above stages.

Overall, the objective of the workshops is to elicit value judgements and the preferences of stakeholders and experts by mimicking the deliberation process used in HTA decision-making; the selected criteria are validated, options are scored against the criteria, and criteria are weighted according to their relative importance, producing overall value scores for each option so that they can be ranked according to their total value.

Results are analysed using tables and graphical illustrations where each individual criterion contribution to overall score is revealed; plotting the scores of any criterion or group of criteria against others (e.g. therapeutic vs. safety) permits one to compare options and visualise their performance against specific value dimensions. Sensitivity analysis is conducted to test the impact of different possible weights or scores on the overall value of the alternative treatments, thus testing input vagueness and possibly helping to resolve any existing preferences disagreements between the various stakeholders. The outcome is a combined perception of total value with the level of information available up to that point.

From a policy perspective these two papers provide case studies of applying in practice a conceptually new methodological MCDA framework for assessing the value of new medicines for the treatment indication of cancer, while empirically producing quantitative evidence relating to their value, in the form of MCDA weighted value scores.

In order to undertake and complete these tasks, effort is needed in applying decision science methods and techniques in practice, collecting both secondary and primary data, conducting literature reviews, presenting the results and consulting with experts, organising, conducting and facilitating decision conferences, combining evidence together and analysing results, and producing policy recommendations.

1.2.3 Research Design and Methodology

Overall this thesis adopts a deductive logic approach according to which an idea is developed (development of the new value framework) and then evidence is collected to test this framework (application of the value framework) following which a discussion takes place

⁹⁶. Chapter 4 (Paper 1) adopts an exploratory research design investigating what are some of the critical features for a robust methodological approach of value assessment for NMTs in the context of HTA and based on MCDA principles⁹⁶. Chapter 5 (Paper 2) adopts a descriptive research design outlining what evaluation criteria are considered as part of current HTA methods in different countries, analysing existing data through a systematic literature but also generating primary data through expert consultations⁹⁶. Chapter 6 (Paper 3) adopts a combination of exploratory and descriptive research designs, making use of the findings from the previous two chapters in combination with decision analysis theory to explore and outline the development of the new value framework. Chapters 7 and 8 (Papers 4 and 5) adopt an evaluative research design to test the newly developed value framework in practice by aiming to assess the value of a set of alternative therapies and generating new primary data through two case studies that involve decision conferences for the collection of preferences from participants. Overall, this thesis uses characteristics of qualitative (Chapters 4-6) and quantitative (Chapters 7-8) methodologies, meanwhile adopting participatory/action methodology characteristics⁹⁷. Ultimately, it aims to advance an action for change, possibly creating debate and discussion, being practical and collaborative in nature, while expecting that other stakeholders might engage as active collaborators ⁹⁷.

1.2.4 Policy Implications

The proposed research agenda acknowledges that pharmaceutical policies must often balance multiple conflicting interests. An evaluation framework for medical technologies should satisfy primarily health policy goals and to some extent may afford to accommodate industrial policy goals, by assessing value of innovative technologies according to a comprehensive set of criteria and incentivising investment in socially desirable R&D.

The ultimate goal of pharmaceutical policies in a health system is to promote population health by improving patient health outcomes as efficiently as possible. Pharmaceutical innovation is one key driver for the production of therapeutic benefits in the form of innovative therapies. Both a comprehensive value-based assessment system and well-aligned incentives are necessary in order to encourage R&D and sustain future innovation. Only in this way will resources be allocated in the most efficient way, maximising social welfare through the development of technologies with the greatest value. At the same time, health systems need to contain escalating pharmaceutical expenditure and

ensure that patients continue to have access to such innovative technologies with high (social) value.

Decision-makers need clear and explicit criteria in order to make transparent decisions about value which can then enable them to make health care coverage decisions based on complete and robust information that reduces risk and mitigates uncertainty on value, while at the same time providing them with stability and predictability in budget allocations. This thesis aims to provide decision-makers with additional tools to make these decisions. The key principle of VBP is to ensure that public health resources are used in a way that produce the greatest possible value for patients. However, methodological and conceptual pitfalls of existing VBP methods mainly relating to the incomplete and *ad hoc* assessment of key value dimensions act as an obstacle. This thesis will have direct applicability to health care decision-makers seeking to make coverage decisions for new technologies and the extent to which these are included in the benefits catalogue.

1.3 Conclusion

The methodological approach for allocating health care resources in an efficient and fair way that gives legitimacy to the decision outcomes has been far from clear. The British National Health Service provides an insightful example where priority setting takes place across all hierarchical levels of the health care system from central government, which decides on the overall budget, to commissioners which determine purchases among alternative interventions, to clinicians who allocate their resources. Over the last two decades, the decision context has remained similar, together with the use of the QALY that has acted as the decision-making metric for budget distribution ¹¹. However since then, research has identified a number of methodological challenges due to technical and problematic concepts that would have to be overcome for the decision-making process to improve and outcomes to become rational. For example, Klein had suggested the exploration of conflicts between different values or preferences and testing of arguments against evidence as part of open dialogues ¹¹, whereas Daniels' and Sabin's Accountability for Reasonableness ethical framework has been proposed as a way of producing fair decisions through a legitimate process where they are publicly available, based on relevant reasoning and revisable in light of new evidence ²⁰.

Pharmaceutical costs are the third largest component of health care expenditure in most OECD countries, experiencing an escalating trend that is partially caused due to the uptake of new branded medicines with high prices in a small number of therapeutic

categories such as cancer treatments. Clinical uncertainty of new medicines in combination with their high costs has led to the evaluation of their benefits and costs by payers as part of HTA, in order to estimate their value, coverage and possibly price. The rewarding mechanism for innovative medicines has to consider both the current value to consumers, *i.e.* patients, while also encouraging future societal gains, with the emergence of value based assessment and pricing suggested as a way of improving patients' access while ensuring that prices reflect value to a variety of stakeholders.

A mix of objectives would have to be balanced out, possibly though the application of a plurality of evaluation criteria characterising the intervention of the target indication, however most priority setting approaches tend to focus on single value dimensions despite the fact that the value of new medicines is multi-dimensional in nature. Current value assessment approaches tend to work towards comparative effectiveness and efficiency, but there is considerable arbitrariness in the selection of evaluation criteria and how to account for other key benefit dimensions. Explicit elicitation of social value trade-offs is not possible and decision-makers may adopt intuitive or heuristic modes for simplification purposes given these multi-criteria problems, based on *ad hoc* decision making procedures. A more comprehensive and transparent methodological approach incorporating a wider set of benefits and their potential value trade-offs based on stakeholder preferences, may lead to the development of a decision support tool for decision-makers that contributes towards a more efficient, rational and legitimate resource allocation.

The research question of the thesis is whether and how a methodological framework can be developed in order to capture value of NMTs in a more encompassing way. By departing from current principles of value assessment through traditional HTA, the thesis argues in favour of the use of MCDA in the context of HTA, rather than economic evaluation.

In order to fulfil the research objective, five research papers are undertaken, addressing each of the secondary research questions separately and comprising Chapters 4 to 8 of this thesis as following. Chapter 4 (Paper 1) forms the conceptual framework of the thesis and outlines the methodological process adopted for the development of the MCDA value framework, addressing research question 1. Chapter 5 (Paper 2) is a systematic literature review and expert consultation in Europe, addressing research question 2, which feeds the development of the next paper. Chapter 6 (Paper 3) forms the methodological framework output of the thesis and describes the development of a value measurement model, addressing research question 3. Chapters 7 and 8 (Papers 4 and 5) form real world

applications of the newly developed value framework, addressing research question 4 in the form of two separate case studies. Chapter 9 outlines the conclusions of this thesis, examines the policy implications of the findings, outlines the limitations of the methods and results, and recommends future research directions.

In sum, the intellectual and academic contribution of this thesis progresses in parallel with the subsidiary research questions and relates to the research, design, development and testing of a new value framework in practice for the purpose of HTA, based on MCDA principles and in alignment with decision theory.

Chapter 2 – Current Methods of Value Assessment of Medical Technologies

As part of current HTA practices, a number of value assessment approaches are used across different settings that can be applied on their own or in combination, namely (a) assessment of clinical benefit, (b) economic evaluation and (c) risk sharing and managed entry agreements. But are these methods ‘sufficient’ to capture the overall value of NMTs and aid the decision-making process? If not, what other methods might act better in assessing value and why?

2.1 Current Methods of Value Assessment in Health Technology Assessment

The concept of HTA was originally applied in 1978 by the US Office of Technology Assessment ⁹⁸. HTA can be defined as “a method of evidence synthesis that considers evidence regarding clinical effectiveness, safety, cost-effectiveness and, when broadly applied, includes social, ethical, and legal aspects of the use of health technologies”, with the aim of advising on coverage and reimbursement decisions ⁹⁹ (p. 271). Currently, most countries in the European Union (EU) adopt some form of HTA procedure as a means of assessing the value of NMTs for health care systems with publicly funded health services ¹⁰⁰. The overall purpose of HTA is to determine the value of a health technology by evaluating its benefits and - and most often - its costs, compared to existing alternatives of care, by considering both clinical and economic evidence in order to inform coverage and pricing decisions ^{101,102}, predominantly of new medicines ⁹⁹.

2.1.1 Assessment of Clinical Benefit

The simplest type of HTA centres around the evaluation of the clinical/therapeutic benefit of a new drug relative to a comparator, usually the best available alternative or the standard of care. This is a criterion applied by the vast majority of EU countries, either on its own (e.g. France and Germany) or in combination with other criteria (e.g. through cost-effectiveness in the United Kingdom or Sweden) as part of assessing the value of NMTs.

By comparing the relative therapeutic benefit across different medicines, their marginal clinical value is assessed and can be linked with reimbursement and/or pricing decisions ²⁸. The higher the relative therapeutic benefit of a product, the higher its clinical value, hence the stronger the justification for coverage and premium pricing.

In most cases this therapeutic or clinical benefit relates to the clinical performance of the drug in patients that act as the subjects of RCTs, that are conducted as part of the marketing authorisation process, thus relating to efficacy ¹⁰³. Ideally however, the clinical performance evaluated should be based on evidence from real world clinical practice that reflects effectiveness.

Although a suitable comparator should be the most relevant competing medicinal product currently available for the respective disease indication ¹⁰⁴, in reality this is rarely the case because of the lack of head to head clinical trials. Instead, most clinical trials use placebo or best supportive care as a comparator. Consequently, in order for the therapeutic benefit of different drugs to be assessed against each other, an indirect or mixed comparison approach is followed ^{105,106}.

2.1.2 Economic Evaluation

Once the additional criterion of cost is incorporated as part of a more comprehensive value assessment process, then the procedure starts to follow the concept of economic evaluation. Different HTA approaches use the basic concept of traditional economic evaluation, or economic analysis, according to which the costs and consequences of an alternative A, are compared with the cost and consequences of another alternative B. More precisely, the difference in costs is compared with the difference in outcomes (or effects) for two or more alternatives and the cost per unit of outcome is assessed, known as incremental cost effectiveness ratio (ICER) which reflects opportunity cost.

Economic evaluation could thus be defined as the comparative analysis of alternative courses of action in terms of both their costs and their consequences¹⁰⁷, with its basic tasks being to identify, measure, value, and compare the consequences and costs of the alternatives being assessed. In other words, economic evaluations help us to answer the question of whether a particular health technology is “worth doing compared with other things we could do with these same resources”, or similarly whether we are “satisfied that the health care resources should be spent in this way rather than some other way” (p. 7)¹⁰⁸.

The theoretical background of economic evaluation is grounded in welfare economics, a branch of economics providing the basis for evaluating efficiency in resource allocation by markets and policy makers, by developing propositions which may be used to rank (e.g. from better to worse) economic situations based on their associated social welfare^{109,110}. According to welfare analysis theory, individuals aim to maximise their personal utility with societal welfare being the overall welfare of society, equal to the aggregation of these individual utilities or individual welfare; then, the only objective of governments in taking societal decisions is to maximise welfare^{111,112}.

According to the New Welfare Economics approach⁴, social welfare is dependent on the dimensions of efficiency and social justice (equity), and can only be maximised by the joint pursuit of both^{113,114}. Pareto efficiency usually acts as the efficiency goal; this is an aim that is relevant to all theories of society and is a necessary but not sufficient condition for an increase in social welfare. On the other hand, equity or distribution goals will depend on the distributive justice theory adopted (i.e. the definition of social justice chosen), ranging from utilitarian (Benthamite) to Max-Min (Rawlsian) and in between (egalitarian), affecting the specification of social welfare function (SWF) which incorporates value judgments about interpersonal utility¹¹⁴.

Under ideal conditions describing a first-best economy, institutions such as perfect markets will allocate resources optimally (Pareto-efficient) in a manner that agrees with the aim of societal welfare¹¹⁴. However, when these conditions fail and such idealised institutions are absent, intervention is needed and economic evaluation can in theory be used for the efficient allocation of resources^{111,114}. This can take place in the form of cost-benefit analysis, a specific application of welfare economics techniques.

⁴ This approach makes no assumptions regarding interpersonal comparability of utility, in contrast to the Old Neoclassical approach, which starts out with such assumptions.

In the context of health care, the concept of efficiency faces significant difficulties in quantifying benefits; health is hard to measure and health outcomes are often hard to evaluate, causality of improvement in health is complex (e.g. to what extent is improvement in health caused by medical care and not due to natural recovery?), and improved health is difficult to value (i.e. valuing human life could be regarded as immoral and is practically challenging) ¹¹⁴.

There are three main types of economic evaluation, their key difference lying in the nature of outcomes, or effects, being considered (*Appendix* to Chapter 2) notably: (a) cost-benefit analysis (CBA), which measures health outcomes in monetary units; (b) cost-effectiveness analysis (CEA), which measures a clinical effect that is common between the alternatives being compared (e.g. 'life years gained'); and (c) cost-utility analysis (CUA) which measures the utility associated with the effects of comparators being assessed. In the context of CUA, utility refers to the preference of individuals or society on any particular type of health outcomes and is usually measured using the QALY, which as defined above is a utility-adjusted health outcome that combines the dimensions of length of life and quality of life in a single index number. The use of monetary units for the purpose of measuring costs is common across all three types of economic evaluation, however the types of costs included varies by each setting, depending on the perspective adopted.

The selection of which type of economic evaluation to use depends primarily on the nature of the problem being assessed, but also on the practical challenges associated with the measurement of outcomes and costs, the institutional framework, and the perspective adopted in terms of the role of economic evaluation itself ¹¹⁵. Due to limitations associated with the use of monetary measures of benefit⁵ ^{111,116-118}, in combination with a rising availability of new clinical effectiveness measures, the extension to CEA and consequently to CUA is relative easy and straightforward thus turning them into the preferred methodologies ¹¹¹.

However, evidence on the effectiveness of medical treatment is still limited and the QALY approach as we will see below is associated with a number of limitations. Therefore, although the efficiency aim in health care is obvious enough, methodological problems associated with the measurement of benefits renders the assessment of efficiency problematic.

⁵ Assignment of monetary units to health care outcomes is conceived as a difficult task and willingness-to-pay (WTP) may be influenced by ability to pay thus possibly reflecting non-health attributes of health care options.

2.1.3 Risk Sharing and Managed Entry Agreements

Within the context of HTA, managed entry agreements (MEAs) act as an additional tool aiding decision-makers to account for uncertainty related to clinical outcomes or the overall budget in a particular indication ²². The term MEA is collectively referring to a variety of schemes for addressing clinical uncertainty due to lack of information on effectiveness at the time of HTA appraisal, such as risk-sharing agreements and performance-based agreements, with the aim of sharing the financial risk between payers and manufacturers ¹¹⁹. For instance, through the use of MEAs, coverage may be gained once health outcomes uncertainty has been addressed; this can mean that the use of a new treatment is restricted to specific sub-groups of patients with satisfactory clinical performance in relation to cost, or through the submission (and acceptance) of additional, real world evidence of the treatment's effectiveness by the manufacturer often in combination with price discounts and money-back guarantees ¹²⁰. The usefulness of MEAs in policy-making relates to the development of effectiveness evidence in order to reduce risk and uncertainty about the value of a new technology; yet, such additional evidence do not necessarily take into consideration a broader set of value parameters.

2.2 Limitations of Current Value Assessment Approaches

The way value of NMTs is conducted across different settings presents a number of shortcomings, in particular (a) the absence of an explicit value for money definition, (b) the inadequate and *ad hoc* nature of the value criteria used and (c) the heterogeneity over HTA recommendations across settings, partly as a result of the above.

2.2.1 Absence of an Explicit Value Definition

As part of CUA, the QALY has been established as the preferred type of outcome many HTA bodies use, thus acting as a central variable of interest ¹²¹. Although EQ-5D is a generic instrument used to derive patient quality of life across different health states, there is a debate of whether it reflects patient preferences accurately enough in order to study a treatment's impact, given that reference scores used to build the measuring scale have been elicited from healthy individuals ^{122,123}.

By comparing the ICER of different alternatives using cost per QALY gained, an option with lower ICER is preferred over an alternative with higher ICER. An external

reference point is usually applied that reflects a ‘value for money’ range of acceptance, usually taking the form of a willingness to pay (WTP) threshold from a decision-maker’s perspective such as WTP for the gain of an additional QALY. In England, the National Institute for Health and Care Excellence (NICE) is the HTA agency responsible for producing evidence based guidance and advice for health care resources with the ultimate goal to improve outcomes for people using the NHS and other public health and social care services. As Devlin and Parkin had acknowledged ¹²⁴, NICE had stated in its own guidance that a “sufficient level of cost-effectiveness is in the range of £20,000 and £30,000” ¹²⁵, however NICE later reported that such comments were misinterpreted and that it does not have such a threshold ¹²⁶; it is therefore believed that the above range constitutes an implicit range ¹²⁷⁻¹²⁹. A range of different WTP thresholds exist across different HTA settings ¹³⁰, which are however arbitrary in nature and lack a scientific basis of elicitation.

Evidence support that NICE coverage decision outcomes are affected by several factors beyond cost-effectiveness estimates (including uncertainty, burden of illness, quality of the evidence, availability of treatments and budget impact) and that modelling three decision outcomes (i.e. routine use, restricted use, no recommendation) instead of two (i.e. positive, negative decisions) better explains decision processes in some settings ¹³¹. Other research suggests that for drugs of particular therapeutic areas such as cancer and orphan diseases, any cited implicit thresholds may not be rigidly adhered to, showcasing the recommendation of drugs with ICER up to £59,000 by NICE and other agencies, and proposing the existence of differential weights of importance across different diseases ¹³²⁻¹³⁶.

2.2.2 Value Criteria: How Complete Are They?

Besides important theoretical and conceptual limitations inherent in the calculation and use of QALYs in health care decision-making ^{16,17,137,138}, the argument that the incremental ‘cost per QALY’ metric does neither adequately nor explicitly capture a number of important value dimensions suggests that it cannot be used as a sole approach to value assessment ^{28,132,139}. Additional criticisms of current HTA processes include the relative lack of transparency and existence of uncertainty ^{133,140-144}. Altogether this evidence is consistent with other studies indicating the rationale for a shift towards more transparent VBP ^{26,27,29,30,145}.

There is generally an acceptance of the fact that other social factors in addition to ‘scientific value evidence’, based solely on clinical cost-effectiveness and the ICER, should play a role in decision-making. These include burden of disease, the broad balance between benefits and costs, cost impact on available budgets and resources, clinical and policy importance of the health topic under consideration, and other policy objectives such as the long-term benefits of innovation ¹⁴⁶⁻¹⁴⁸, including the balancing of long-term product innovation (dynamic efficiency) with short-term product pricing (static efficiency) ^{149,150}.

However, it is not known which of the above ‘social value judgments’ are explicitly or transparently considered by HTAs or decision-makers, what their relative importance is, and what the trade-offs are that HTA bodies are willing to make. This gives rise to arbitrariness in the set of criteria used, resulting in non-transparent processes and lack of ‘accountability for reasonableness’ ^{151,152}; in turn, important equity and fairness considerations can emerge relating to the distribution of health care among individuals, including the conditions under which it is acceptable to place more weight on a QALY for one person compared to another ¹⁴⁸. The importance of special considerations such as disease severity is illustrated in the case of end of life treatments in the UK context, in essence giving a greater weight on QALYs gained for terminal illnesses ¹⁵³, on the grounds that society places a special value on extending the lives of the terminally ill ⁶⁷. The on-going debate in the UK about VBP reflects additional considerations that need to be taken into account in order to help determine the full value of a health technology based on a broader range of relevant factors ^{31,139,154}.

2.2.3 Coverage Heterogeneity

The above limitations associated with the arbitrary and implicit definition of value and the subjectivity in the evaluation criteria used to measure it, often lead to differences in the assessment and appraisal processes resulting in discrepancies of health care coverage recommendations across settings, even for the same drug-indication pair ^{101,155-160}. Other reasons which may affect the impact and outcome of HTA recommendations include differences in responsibilities, abilities and national priorities of the HTA agencies and the constraints and timeframes in which they operate ¹⁶⁰⁻¹⁶⁴.

Decision outcomes and coverage heterogeneity can be justified on the grounds of different national (or, at times, regional) preferences, including different budget constraints. However, inconsistencies of medicines’ eligibilities for reimbursement across geographical

jurisdictions, possibly even within the same country, give rise to an international ‘post-code’ lottery type of access to medicines, with important implications related to equity and fairness¹⁰¹. Indeed, several studies have acknowledged the need for well-defined decision-making procedures that are more fair and explicit¹⁶⁵⁻¹⁶⁷. By ensuring ‘accountability for reasonableness’ and by providing a better understanding of the rationale behind decision-making, decisions will also have enhanced legitimacy and acceptability^{152,155}.

2.3 Summary of Gaps in the Literature and Likely Ways Forward

2.3.1 Gaps in the Literature

Most of the current value assessment approaches seem to examine the efficacy/effectiveness, or cost-effectiveness of new health care interventions by addressing only a partial dimension of ‘overall value’ that mainly relates to ‘scientific value judgements’ of their therapeutic aspect (e.g. safety, efficacy, effectiveness), possibly in relation to cost. For example, the cost/QALY approach considers the incremental (quality-adjusted) therapeutic benefit, in the form of QALYs gained, relative to its incremental cost. Hence, current methodologies have been subjected to criticism because they neither adequately nor explicitly capture social value judgement elements related to the burden of disease the treatment addresses, the overall innovation level of the treatment, and wider socioeconomic implication elements. Therefore, in the context of value assessment for NMTs, the following gaps have been identified:

1. A conceptual framework for value definition is needed that takes into account a wider set of explicit evaluation criteria.
2. A clear, transparent and comprehensive set of evaluation criteria to be used in value assessment is also required, pointing towards the need for a methodological framework.
3. Considering the limitations of the cost/QALY metric, few other, if any, approaches have been put forward that capture the link between value and HTA decisions in order to achieve efficient allocation of resources.

More robust and complete assessment procedures characterised by transparency in terms of value criteria used, possibly taking the form of value functions, could lead to more rational evidence-based decision-making, hence to more efficient resource allocation and higher society welfare, while also raising public confidence and fairness in terms of homogeneity and consistency of decision outcomes.

2.3.2 Decision Analysis and Multiple Criteria Decision Analysis as a Means of Addressing the Pitfalls of Current Value Assessment Methods

The limitations of the current approaches in value assessment and the identified literature gaps suggest that there is a need for an alternative methodological approach of value assessment that encompasses multiple criteria explicitly, so that value can be a function of a number of parameters. Additionally, questions remain on how to derive weights for the different factors and how to incorporate the views of all relevant stakeholders.

Decision analysis (DA) could provide the foundation for an alternative way of measuring and eliciting value of new therapies as it provides a comprehensive approach for quantitative modelling ⁵³.

More specifically, analysis of the literature on methodological tools for assessing value quantitatively as part of decision-making process indicates MCDA; it acts both as an approach and a set of techniques, ordering a set of options by looking at the degree to which a number of different objectives are achieved ^{1,168}. It is a way of eliciting preferences for a sum of options which are characterised by varying levels of performance with respect to a number of, often conflicting, objectives; it does so by disaggregating a complex problem into simpler components or objectives, measuring the performance of options against the objectives, weighting up these objectives according to their relative importance, and re-assembling the components by aggregating scores and weights to show the overall picture ^{1,53}.

One of the key aims of MCDA techniques is to enable decision-makers to reach a decision by laying out the problem, objectives, values and options they are faced with in a clear and transparent way. This is achieved by organising, synthesising and presenting relevant information across stakeholders, which is of complex and, often, of conflicting nature ². Although results will always be as good and certain as evidence used, “uncertainty on value” could be improved because of a more structured and comprehensive assessment based on wider value dimensions. While MCDA can aid the decision-making process, it cannot replace decision-makers’ judgement or experience ². However, it can supply to decision-makers detailed information for a comprehensive set of parameters of interest, in a transparent way and across a wide range of stakeholders at the same time. Overall, MCDA acts as an aid to decision-making, seeking to explicitly integrate objective measurement with value judgement while managing subjectivity in a transparent way.

2.3.3 Emergence of New Value Frameworks Based on Multi-Criteria Approaches

Over the past few years, the need for a more “rational” approach to value assessment and decision-making has been illustrated through a number of initiatives on the development of new value frameworks aiming to aid reimbursement agencies, health care professionals and patients to make better choices on the use of new therapies by better understanding their value ¹⁶⁹. This serves as a testament to the above gaps in the value assessment of NMTs presented earlier in this chapter.

Among some of the most known value frameworks that have attracted a lot of attention include those proposed by the American College of Cardiology and the American Heart Association (ACC/AHA) ¹⁷⁰, the American Society of Clinical Oncology (ASCO) ¹⁷¹, the European Society of Medical Oncology (ESMO) ¹⁷², the Institute for Clinical and Economic Review (ICER) ¹⁷³, the Memorial Sloan Kettering Cancer Centre (MSKCC) ¹⁷⁴, the National Comprehensive Cancer Network (NCCN) ¹⁷⁵, and the Working Group on Mechanisms of Coordinated Access to Orphan Medicinal Products (MoCA-OMP) ¹⁷⁶. Essentially these value frameworks adopt multi-criteria evaluation approaches in an attempt to decompose complex problems into simpler ones and address them separately, therefore representing an important step towards a more inclusive Value Based Assessment (VBA) process. However aspects of these frameworks are based on weak methodologies with no strong theoretical foundations and as such might offer little value to decision-makers ¹⁷⁷. Despite this proliferation of value frameworks, “value” remains an elusive target and a wider consensus is still needed about what dimensions of “value” to be included in HTA ⁶⁹.

2.4 Conclusion

Most EU countries adopt some type of HTA procedure in order to assess the value of NMTs and more precisely that of new medicines in order to inform coverage and pricing decisions. These procedures include the assessment of the clinical benefit, assessment of clinical benefit in combination with related costs as part of different types of economic evaluation, and risk sharing and managed entry agreements aiming to reduce risk and uncertainty by delaying reimbursement decisions until data on real clinical outcomes (i.e. effectiveness) and costs become available.

However, technical shortcoming due to methodological and theoretical limitations in the application of economic evaluation such as the absence of an explicit value definition and the use of incomplete and often arbitrary value criteria can result in *ad hoc* evaluation procedures that give rise to coverage recommendations, often lacking any ‘accountability for reasonableness’ that might negatively influence their legitimacy and acceptability.

It seems that most of the current value assessment approaches address only partial dimensions of medicines’ overall value that mainly reflect scientific value judgements of their therapeutic benefit possibly in relation to cost, without however adequately nor explicitly capturing elements of social value judgements. As a result, in the context of value assessment there is a lack of a conceptual and methodological value framework that takes into account a wider set of evaluation criteria in a transparent and comprehensive way with few alternative methodologies to economic evaluation having been tested.

An alternative methodological approach of value assessment that encompasses multiple parameters explicitly seems to be needed, and MCDA could act as a way of eliciting preferences for a set of NMTs which are characterised by varying levels of performance with respect to a number of, often conflicting, objectives. Overall, such a methodology could possibly act as a decision-making tool, seeking to explicitly integrate objective measurement with value judgement while managing subjectivity in a transparent way.

Chapter 3 – Paper 1

Conceptual Framework: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment⁶

Summary

In recent years, multiple criteria decision analysis (MCDA) has emerged as a likely alternative to address shortcomings in health technology assessment (HTA) by offering a more comprehensive perspective to value assessment and acting as an alternative priority setting tool. In this paper, I argue that MCDA needs to subscribe to robust methodological processes related to the selection of objectives, criteria and attributes in order to be meaningful in the context of healthcare decision-making and fulfil its role in value-based assessment (VBA). I propose a methodological process, based on multi-attribute value theory (MAVT) methods comprising five distinct phases, outline the stages involved in each phase and discuss their relevance in the HTA process. Importantly, criteria and attributes need to satisfy a set of desired properties, otherwise the outcome of the analysis can produce spurious results and misleading recommendations. The application of MCDA as described here has the potential to offer three distinct advantages to decision-makers in the context of HTA and VBA. These relate to, first, a more complete assessment of value arising from the elicitation of preferences across a wider set of explicit criteria, second, the flexibility in distinguishing value through the incorporation of weights that reflect differences in relative importance, and, third, the encompassing nature of the entire process which can be informed by direct stakeholder engagement, all features being fully transparent and facilitating decision-making.

3.1 Introduction

The use of economic evaluation methods, particularly cost- effectiveness analysis (CEA) and cost-utility analysis (CUA), to assess the incremental benefit of new medical

⁶ Chapter 3 has been published with co-author Prof Panos Kanavos as: Angelis A, Kanavos P (2016). Value-based assessment of new medical technologies: towards a robust methodological framework for the application of multiple criteria decision analysis in the context of health technology assessment. *PharmacoEconomics*. May;34(5):435-46.

technologies in relation to the best alternative care has increased considerably over the past 2 decades. In this context, the use of the QALY has been established as the preferred measure of health gain across many settings ^{108,178-180}. This is despite its frequent dependence on restrictive assumptions ¹⁸¹, the non-alignment of public versus patients' decision utilities, which would differ from their respective experienced utilities ¹⁸², and the reliance on generic tools, such as the EQ-5D, that may not reflect patient experience adequately ¹⁸³⁻¹⁸⁵.

At the same time, there is increased recognition that economic evaluation has limitations because it does not capture a number of important dimensions of value, and is therefore lacking in comprehensiveness. In partial recognition of that, economic evaluation has recently evolved into a deliberative process across different settings, whereby independent decision-making committees often allow for other dimensions of value to be considered, at least implicitly.

Additionally, there is increasing evidence that decision-makers are reluctant to make coverage recommendations on economic evaluation alone ¹⁸⁶ and, consequently, 'value' based on economic evaluation results could be informed by additional dimensions of benefit. Recently, decision-makers in England and Wales considered additional parameters of benefit on an *ad hoc* basis ¹⁸⁷, highlighting the need to seek a broader and more transparent assessment methodology ^{31,154}, in the context of value-based pricing ^{28,30,188,189}.

Even under such enhanced settings, the decision-making framework often lacks transparency, not least because different stakeholders attach different value judgements to the criteria considered. Consequently, value assessment is not simply a question of what additional benefits to consider and possibly include in the decision-making process, but, importantly, involves how to arrive at a clear process that elicits and accounts for the preferences of different stakeholders in a transparent way. The ongoing debate in the UK on value-based pricing is a testament to these issues ¹⁹⁰.

Overall, the lack of comprehensiveness in value dimensions considered, the non-systematic use of additional value concerns and the lack of transparency in terms of decision-makers' value trade-offs, altogether reflect an arbitrariness and inconsistency in the appraisal process. As a result, 'unexplained' heterogeneity or inconsistency in the decision-making process could have important implications for fairness, equity and resource allocation outcomes. Therefore, the development of alternative methodological approaches for assessing the value assessment of medical technologies that would overcome the above limitations could eventually contribute towards a more efficient resource allocation with improved impact on equity that would be perceived fairer by society.

In recent years, multiple criteria decision analysis (MCDA) has emerged as a likely alternative to address the current shortcomings of Health Technology Assessment (HTA) based on economic evaluation ^{90,91,191-194}. One of the conclusions of a recent review of MCDA approaches adopted in healthcare was that decision-makers display a positive attitude towards its potential to improve decision-making ⁷². Conceptually, there are three main reasons why MCDA could provide a useful alternative to economic evaluation-based HTA processes. The first relates to the inclusion of a comprehensive list of value dimensions in an explicit manner, beyond what economic evaluation methods currently capture. This enables value assessment to be conducted in an encompassing manner and, in principle, addresses a key limitation of economic evaluation. The second relates to the assignment of quantitative weights across the different evaluation criteria. In doing so, the relative importance of various value dimensions is explicitly incorporated, improving the transparency of the preference-elicitation process. The third is stakeholder participation and the possibility to include all relevant stakeholders in the value-assessment process. This is both insightful - enabling stakeholder views to be heard in a dynamic environment, where all inputs are considered prior to making decisions about coverage - and politically correct, increasing the legitimacy of decision processes, as all stakeholder views are accounted for in an open and transparent way.

Despite the above, the methodological details of MCDA implementation in the context of healthcare decision-making have not been sufficiently discussed, and there is no adequate guidance on how MCDA should be conducted in HTA, particularly in relation to which criteria to incorporate and how.

In this paper, I outline a methodological process for the development of a robust MCDA framework and debate its implementation in the context of HTA. In doing so, I provide a broad classification of MCDA methods while also accounting for and building on the classifications proposed in the literature ^{1,2,54,92,195-197}. I then focus on value-based methods, specifically MAVT methods, and argue in favour of using these because of their comprehensive nature. Further, I argue that several key principles need to be fulfilled in order for any MCDA framework to be methodologically sound and for the results produced to be robust and policy relevant. These principles apply to the MCDA main phases and stages as well as to the properties that the selected criteria and attributes need to satisfy, while establishing their relevance in the context of HTA and value-based assessment (VBA). I discuss these principles and their implications in the context of HTA by drawing on concrete

examples. Finally, I discuss a number of practical issues relating to the use of MCDA in HTA and provide a link to policy making.

3.2 A Methodological Process Applying MCDA Principles in HTA and Value Assessment of Medical Technologies

3.2.1 Overview of Different MCDA Methods

MCDA is both an approach and a group of techniques aiming to aid decision-making by laying out the problem, objectives and available options in a clear and transparent way. Different MCDA methods exist, with variable degrees of complexity making use of different analytic models. These methods can be broadly categorised by ‘school of thought’, notably (1) value-measurement methods, including (multi-attribute) value theory and utility theory methods, (2) ‘satisficing’ and aspiration level methods, (3) outranking methods, and (4) other methods such as fuzzy and rough sets methods ^{1,2,54,92}.

However, no universal categorisation of MCDA methods exists, and others have proposed groupings that differ from the above ¹⁹⁵⁻¹⁹⁷. Each MCDA method has its own advantages and disadvantages. The choice of method is informed by the type of problem to be addressed, the type of judgements required, the set of axioms employed to support decision-making, and the kind of responses needed. Some methods address choice problems, while others address ranking problems or classification and sorting problems.

Value measurement methods aim to order a set of alternative options through the production of overall numerical value scores. These methods are appropriate when the analysis considers criteria to be compensatory in nature, i.e. when a bad performance on one criterion can be compensated from a good performance on another criterion. They include *multi-attribute value theory* (MAVT) methods for deterministic consequences, and *multi-attribute utility theory* (MAUT) methods (see section 1.1.4).

Most of the outranking methods require similar procedures and data availability to value measurement methods based on which options are selected, their performance in respect to a set of criteria is assessed, weights are assigned and evidence is aggregated together; however, by some it is assumed that it lies out of the MCDA methodological spectrum, being a separate type of methods themselves ¹. Outranking mainly recognises the fact that options with poor performance on just a single criterion could be practically unacceptable in reality, as commonly the case when politics are taken into account ¹⁹⁸. Outranking is a concept similar to that of dominance, stating that for option A to outrank

option B, there should be enough evidence to judge that A is at least as good as B, with no strong evidence to prove it against. In other words, we could say that alternative *a* outranks alternative *b*, given that there is “sufficient credible evidence” to validate that “*a* is at least as good as *b*, taking all criteria into account”² (p.107). The same type of pairwise comparison must then take place for every pair of options being assessed. The outranking approach acknowledges the fact that preference functions are generally imprecise measures and that the emphasis should lie within the strength of evidence for the affirmation that “*a* is at least as good as *b*”, rather than on the strength of preference itself². As a consequence and in contrast to the preference relationships evident in value measurement methods, four preference states might arise: two when an outranking relationship holds true and two when an outranking relationship is absent; a definite preference for one alternative over the other, an indifference between the two alternatives, and in addition to value measurement preference states an “incomparability” between the two alternatives where lack of strongly enough credible evidence leads to neither a definite preference of an alternative nor to indifference. Essentially, this group of methods is more appropriate for the decision context where multiple alternative options are available and the aim is to identify a subset of options that fulfil some minimum requirements of performance, therefore assuming a non-compensatory nature of criteria performance in contrast to value based methods.

The key assumption underlying satisficing and aspiration methods is the existence of decision-makers’ aspiration levels that characterise whether an option’s performance would be acceptable, or in other words satisfactory. The idea is to identify the closest to satisficing a predefined goal of satisfactory level of criteria achievement through the comparison of alternatives². These methods are also non-compensatory, aiming to address complex problems through formal quantitative decision models. However they are more applicable for decision problems where the alternative options are not simultaneously known but instead they become available sequentially, or in situations where there is such a high number of alternatives available that cannot be assessed simultaneously¹⁹⁹.

3.2.2 MCDA in the Context of Multi-Attribute Value Theory

The methodological process I am proposing in this paper for the context of HTA pertains to the category of value-measurement methods. This is predominantly because of the multiple problems that can be addressed, the simplicity of the judgements required and the relatively limited restrictions imposed by the axioms employed. The value-measurement methods

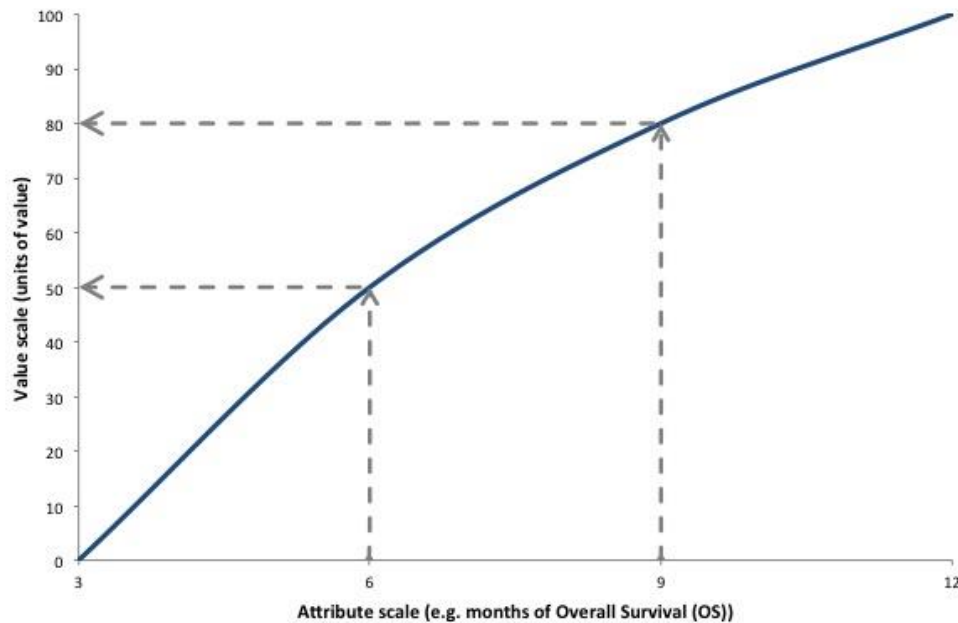
category is widely used in healthcare because of these features. Nevertheless, some aspects (e.g. the MCDA phases and the criteria properties) are applicable across different MCDA methods beyond the value-measurement methods category.

Value-measurement methods usually aim to address ranking or choice problems, ordering a set of alternative options with respect to their performance on a number of objectives or criteria, through the production of overall numerical value scores. A value (or real number) V is associated with the performance of an alternative a , in order to produce an ordering of preferences for all alternatives being considered, while being consistent with the assumptions of complete and transitive preferences. These methods include linear additive methods, multi-attribute value theory (MAVT) methods for deterministic consequences, and multi-attribute utility theory (MAUT) methods.

I suggest the use of MAVT methods because of their comprehensiveness and methodological robustness⁵⁴, as well as their ability to reduce ambiguity and motivational biases. This is in consistency with recent work on good practices for MCDA in Health Care Decisions²⁰⁰. The MAVT methods framework adheres to a number of phases and stages and includes (1) the definition of objectives, (2) the selection of criteria, (3) the scoring of options, and (4) the assignment of weights to the selected criteria.

The choice of technique that will inform parts of the process, including scoring, weighing and aggregation, is an important decision. Under MAVT methods, partial value functions for individual criteria are constructed in the first instance and are subsequently aggregated. Essentially, value functions reflect decision-makers' preferences for different levels of performance on the attribute scale (Figure 4.1). Importantly, the assumptions required for the formation of the partial value functions are interlinked with the aggregation type of technique used. In the sections that follow, I present and discuss these fundamental principles in the context of healthcare decision-making and use examples to illustrate their application and interpretation.

Figure 3.1: Value function for scoring the performance of alternative options



3.2.3 MCDA Phases under MAVT Methods

While the general features of MCDA phases have already been discussed elsewhere ², the MCDA process could be divided into five distinct phases in the context of HTA; these would be (1) problem structuring, (2) model building, (3) model assessment, (4) model appraisal, and (5) development of action plans (Figure 4.2). These phases, and especially model building stages, are taking place in an iterative and constructive way rather than being linearly followed.

Problem structuring involves an understanding of the problem to be addressed. This includes key concerns, envisaged goals, relevant stakeholders that may participate in or contribute to decisions, and identification of uncertainties in terms of a new technology's clinical evidence and its quality.

The phases of model building, model assessment and model appraisal involve the construction of decision-makers' value judgements within and across the criteria of interest, while being consistent with a set of assumptions, aiming to help decision-makers elicit and order their preferences across the alternative options evaluated. For example, if overall survival (OS) is a criterion, then the respective value associated with a range of plausible

incremental OS gains (e.g. 3, 6, 9, or more months) is of interest to know and so is the intensity with which stakeholders would prefer certain changes within the attribute range (e.g. an increase in OS from 3 to 6 months could be of greater value to some stakeholders than an increase in OS from 6 to 9 months) (Figure 4.1).

Finally, given that the outcome of the analysis needs to inform decision-making, action plans need to be shaped involving a clear pathway for result implementation. In the case of HTA, this could involve prioritising resource allocation as part of coverage decisions that take place following the evaluation of new medical technologies.

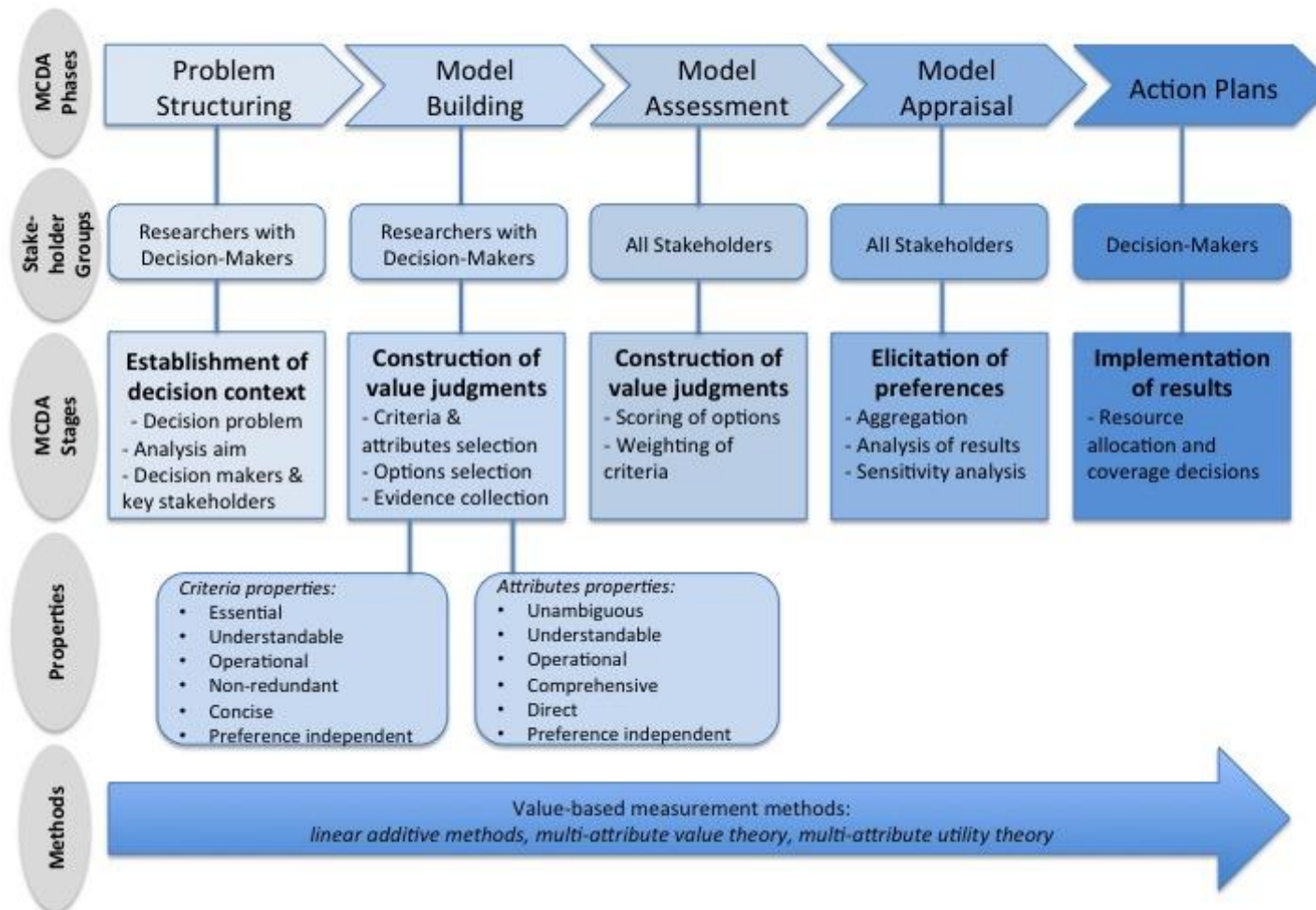
Although these five phases are presented as part of a linear process, in reality they could be part of an iterative process, moving from a later step back to a previous step before advancing. For example, as part of the model assessment phase, it could become evident that some of the criteria do not possess all the required properties, in which case the model should be adapted accordingly as part of the model-building phase.

3.2.4 MCDA Stages under MAVT Methods

3.2.4.1 Problem structuring

Each MCDA phase comprises a number of stages (Figure 4.2). Initially, as part of the problem-structuring phase, the decision context needs to be established where the problem under investigation and the aims of the analysis are clearly outlined and defined, and relevant decision-makers and other key stakeholders are identified. For example, in the context of VBA of a new technology, the decision problem may be to assess the new technology's benefits and costs from a broader societal perspective relative to other therapeutic alternatives to identify the most valuable treatment for a health system. The decision-makers in this context would be payers or insurers (including commissioners of care), whereas healthcare professionals, patients and their carers, technology suppliers and methodology experts, including decision-analysis experts conducting and coordinating the MCDA process, would be the relevant stakeholders. The process of identifying the appropriate decision-makers and stakeholders would be specific to the country or setting. This particular phase could be conducted by researchers or, alternatively, an HTA agency in settings where such an agency exists.

Figure 3.2: Multiple criteria decision analysis (MCDA) methodological process in the context of health technology assessment



3.2.4.2 Model building

Subsequently, as part of the model-building phase, objectives need to be established and/or relevant criteria identified to reflect decision-makers' goals and areas of concern. Additionally, attributes need to be selected to operationalise these criteria and enable their assessment. This involves a deliberative process in order to obtain a good understanding of the decision problem and what decision-makers want to achieve (objectives), through which the values of concern (criteria) will eventually emerge. The assessment takes place based on the selected criteria and attributes. For example, when evaluating a new medical technology relative to an older one, criteria from a number of domains could be selected, such as therapeutic benefit, safety profile, burden of illness, innovation level and socioeconomic impact^{192,201}. In principle, these criteria domains would emerge from decision-makers' values of concern; in practice, they could be identified from the literature in combination with semi-structured interviews with decision-makers. Quality of evidence, mainly relating to relevance and validity of the available evidence, is another crucial parameter that should be considered. This phase could be carried out by MCDA researchers in collaboration with the decision-makers and possibly stakeholders whose value concerns should be considered.

As part of the model-building phase, the alternative options need to be selected, and evidence on their performance across criteria/attributes needs to be identified. For example, the treatment alternatives for a particular disease must be identified and data on expected or observed performance across criteria must be collected, either through secondary research (e.g. from published randomised controlled trial results) or through primary research if data are not available from secondary sources (e.g. clinical or patient opinion). Following the completion of this stage, attribute ranges will be set based on the performance of the alternative treatment options that shall inform the next stages of the process. Depending on the technique used, plausible attribute ranges can be set by taking into account any pre-existing preferences of decision-makers in relation to maximum and minimum allowable performance levels on the different criteria. For example, the OS gains of three different treatments could range from 2 to 12 months, and therefore the respective attribute range should be broad enough to include all these gains (i.e. at least from 2 to 12 months). It could also be the case that the decision-maker is not willing to consider any treatments offering incremental OS gains of less than 3 months; in this instance, the attribute range could be rescaled and adapted to decision-makers' revealed preferences (i.e. to range from a minimum

of 3 months upwards), with the treatment option offering 2 months of OS excluded from the analysis.

3.2.4.3 Model assessment

In the context of the model-assessment phase, the performance of options against the identified criteria must be assessed (i.e. scoring, which delivers intra-criteria information), and criteria must be weighed according to their relative importance (i.e. weighing, which delivers inter-criteria information), revealing preferences for different levels of performance within criteria and across different criteria, respectively. In the case of the OS example, a numerical value score would be assigned to the options being evaluated with regards to their performance on OS gains. As part of MAVT methods, the construction of value functions can take place through different techniques (direct rating, indirect, bisection techniques). All require the definition of attribute reference levels that will form the minimum and maximum points of the value scale. Although the two limits of the attribute range are usually assigned a value of 0 and 100, reflecting the minimum and maximum points of the value scale, respectively, other reference points can also be used. Using the OS example, 3 months could be used as the lower reference level and 12 months as the higher reference level, making up the 0 and 100 points of the value scale, respectively. The attribute performance of the options can then be assessed indirectly through the use of the value functions that will convert their performance into value scores (Figure 4.1). The process of scoring and weighing completes the construction of value judgements.

A critical aspect in the entire process is the relative importance of the different criteria to decision-makers. For this reason, relative weights are assigned to the criteria by directly involving decision-makers and stakeholders. For example, in the case of a new drug–indication pair, the importance of the therapeutic benefits vis-a`-vis an existing therapeutic alternative (e.g. OS gain and quality-of-life improvement) could be found to be twice as important as its safety impact (e.g. adverse events); therefore, the relative weight of the therapeutic cluster of criteria would be twice as high as the product’s safety profile. These weights should only be viewed as scaling constants or trade-off factors, with no algebraic meaning, assigned to enable comparability across criteria to reflect their relative importance.

Methodologically, and contrary to what has been argued elsewhere ^{90,194}, I would argue that it is important for the criteria weights to be derived ex-post following the selection of the alternative treatments and therefore the formation of the attribute ranges, rather than

ex-ante ²⁰². Theoretically, this is tantamount to arguing for MAVT models, where the construction of value functions precedes the criteria weights, rather than for direct rating methods, where weights are first attached, based on an ex-ante derivation, and the options are then scored. Conceptually, my preference for the ex-post derivation of weights is justified by the nature of health technologies and the conditions they treat: the relative importance of different criteria and, therefore, their respective weights are context specific and depend on the performance of the alternative options in a given context. By means of an example, assume that for two treatments (A and B), weights need to be established for the same criteria (OS and hepatotoxicity), measured through ‘number of months gained’ and ‘incidence of hepatotoxicity’, respectively. The weight assigned to each criterion is very likely to be different if treatment A and B range between 1 and 10 months (1,10 months) in OS and from 10 to 11 % (10,11 %) in hepatotoxicity, compared to the scenario that they range between 10 and 11 months (10,11) and from 1 to 10 % (1,10 %), respectively.

3.2.4.4 Model appraisal

As part of the appraisal phase, scores and weights are combined to create a value index (‘aggregating’). The details of this step may differ according to the type of aggregation model used, to include additive or multiplicative value models depending on the level of preference independence present among criteria. Empirical evidence suggests that errors due to the use of additive value aggregation models are in real settings very small and considerably smaller than the errors associated with the wrong aggregation of partial value functions that can possibly result from the incorrect application of more advanced models².

Overall, the individual criteria scores and their respective weights are combined to produce weighted scores and are summed to arrive at an overall value score for each treatment option. In combination with sensitivity analysis, results are examined to determine the robustness of the results obtained. The outcome of this process is a ranking of all treatment options based on their respective value scores. Decision-makers can use this evidence to make resource-allocation decisions. Throughout the MCDA stages, including scoring and weighing, the participating stakeholders are able to interact to exchange views, reach consensus or simply provide their individual preferences ². To that end, they can compare their individual views and preferences, they can aggregate such preferences by voting to reach consensus, or they can share commonly defined modelling and judgement elements after joint discussion.

3.2.5 MCDA Techniques Using MAVT Methods

Several MCDA techniques are available with regards to scoring, weighing and aggregating. These techniques mainly relate to the value judgement and preference-elicitation processes, and the choice of technique depends on the particular type of method adopted ^{1,195-197}. As part of MAVT methods, the value functions based on which options are scored can be constructed using different options: (1) direct rating techniques, (2) indirect techniques, and (3) indifference or bisection techniques ^{2,54}. Direct rating techniques involve decisions around the form of the value function and whether they increase monotonically (highest attribute level is the most preferred), decrease monotonically (lowest attribute level is the most preferred), or range non-monotonically (an intermediate attribute level is the most preferred).

Indirect techniques generally assume a monotonic function and involve a series of questions aiming to uncover decision-makers' preferences by considering differences in the attribute scale and their relation to the value scale. Indifference techniques explore the magnitude of increments in the attribute scale that correspond to equal units in the value (preference) scale. Finally, bisection techniques explore the estimation of points on the attribute scale that serve as midpoints on the value (preference) scale.

I would suggest the use of indirect elicitation techniques because of their comprehensiveness and unbiased nature. This is mainly because decision-makers' intra-criterion preferences are first elicited for the attribute range, producing a transparent valuation relationship across the performance of the options and a value scale. Options are then scored indirectly using the attributes' emerging value functions to convert the performance of the options into value scores, without revealing any information about the identity of the respective options at any point during the process. In the case of decision contexts requiring repeat decisions, indirect MAVT techniques making use of value functions are recommended in order to ensure the efficient and consistent scoring of alternatives as they become available for evaluation ²⁰⁰.

An example of such an indirect technique is MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique), a convenient indirect approach to elicit value functions by only requiring qualitative judgements about the difference of value between different pairs of attribute levels ²⁰³. It uses seven semantic categories to distinguish between the value or attractiveness of different attribute levels, ranging between ‘‘no difference in

value’’ and ‘‘extreme difference in value’’. Overall, it builds a quantitative model of values based on qualitative (verbal) difference judgements, and by analysing judgmental inconsistencies, it facilitates the move from ordinal preference modelling to cardinal preference modelling.

Once criteria have been scored and value functions have been derived, criteria weights can be elicited, usually through a swing weighting technique. Finally, criteria scores and weights are combined, usually through an additive aggregation approach.

3.3 Model-Building and the Construction of a Value Tree: Properties to Ensure a Robust MCDA Model

Model building is one of the most important MCDA phases. Establishing objectives and defining the actual criteria and attributes are critical stages in this context because they form the foundation of MCDA. For the analysis to be robust and, ultimately, meaningful, I outline a number of properties to which criteria and attributes should adhere.

3.3.1 Objectives, Criteria and Attributes in the Context of Model Building

Depending on the decision problem under consideration, the term ‘objective’ or ‘criterion’ may be preferred over the other, both representing key factors that form the basis of the analysis. The main difference between the two is that ‘objectives’ usually reflect a direction of preference, whereas ‘criteria’ do not. Objectives and criteria may be further decomposed into sub-objectives and sub-criteria; structuring all objectives and/or criteria in the form of a tree offers an organised overview of the values under consideration. This is known as a value tree. The quantitative or qualitative performance measures associated with criteria or objectives are known as ‘attributes’. Attributes operationalise the use of criteria and objectives by measuring the extent to which criteria or objectives are achieved. For example, in the context of a new cancer treatment, an objective for decision-makers could be to ‘maximise life expectancy’; ‘overall survival’ could act as a criterion, while ‘median number of months from randomisation to death’ could be the relevant attribute (Figure 4.3).

It is not uncommon for a criterion to require more than a single attribute to be measured adequately. For example, in the case of ‘tolerability’ as part of a new drug’s safety profile, decision-makers could benchmark against the ‘proportion of patients discontinuing the treatment’ as well as the ‘proportion of patients interrupting treatment or reducing the

dose due to adverse events'. Other examples of value tree hierarchies—made up of criteria and attributes—together with their respective data sources are shown in Figure 4.3.

Depending on the type of decision problem, the selection of objectives, criteria and attributes can either precede or follow the identification of the alternative options as part of different methodological approaches (Table 4.1) ^{2,3,204}. In the context of the “value- focused thinking” approach, objectives and criteria are selected prior to specifying or assessing the alternative options, thus being part of a top-down manner for structuring a value tree according to which overall objectives or criteria are decomposed into sub-objectives or sub-criteria ³. Alternatively, in accordance with the more traditional “alternative focused thinking” methodology, a bottom-up approach can be implemented whereby objectives and criteria emerge following the comparison of the options, based on distinguishable attributes that differentiate them ²⁰⁴.

In the context of HTA, a “value-alternative hybrid thinking” logic that contains elements from both methodological approaches could be adopted. Decision-makers could have a generic set of predetermined objectives and criteria reflecting their values of concern in a top-down approach. These could then be adapted for the purposes of the decision-making problem in a bottom-up approach. Thus, the general values of concern would be tailor made in a dynamic manner to better assess the differences of alternative treatments being compared. For example, decision-makers’ concerns could normally include the existence of any contraindications or warnings and precautions associated with a drug for the indicated patient population of interest. However, it is possible for all alternative treatments evaluated for a particular disease to have no contraindications and to have identical minor warnings and precautions for use. These criteria could therefore be excluded from that particular assessment in order to be concise.

Figure 3.3: Value tree hierarchies and data sources using a multiple criteria decision analysis (MCDA) process for value assessment

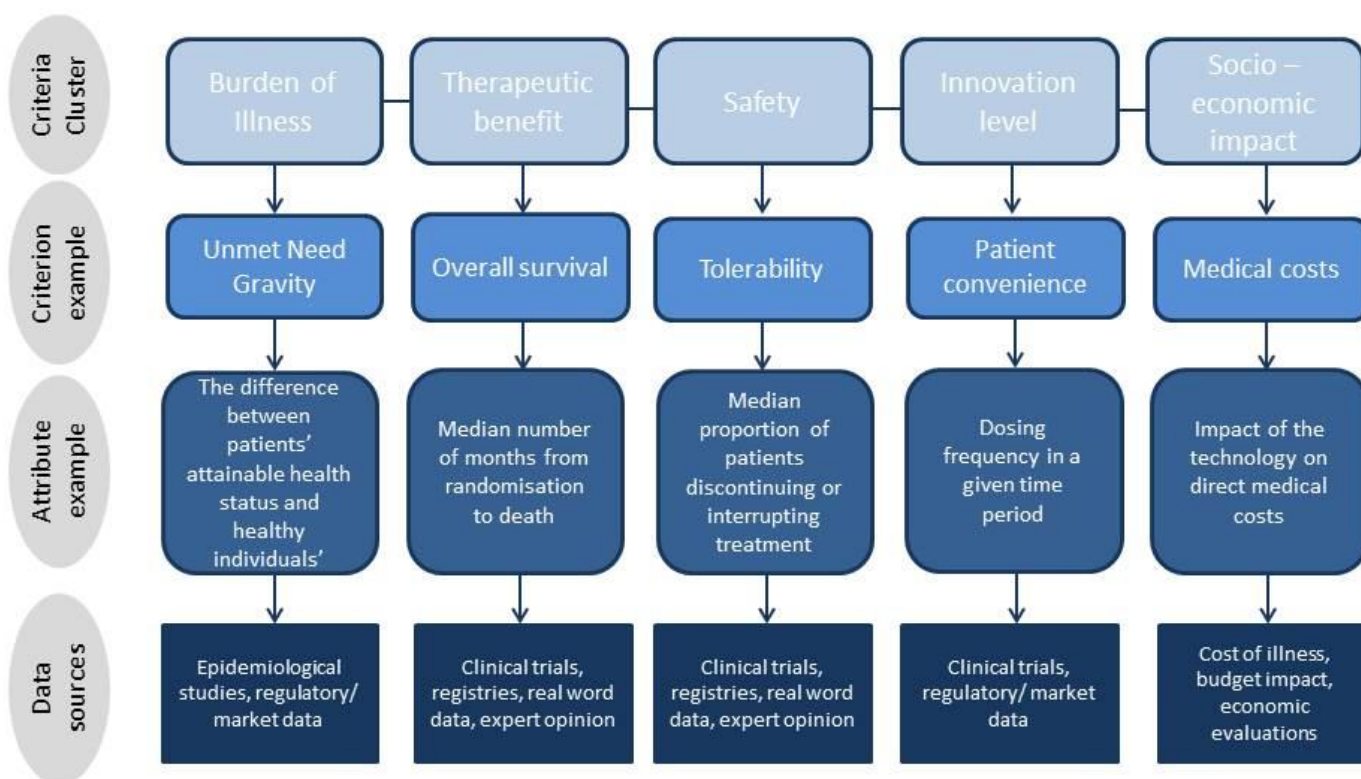


Table 3.1: Different methodological approaches for selecting objectives and criteria

Approach	<i>“Value focused thinking”</i> ³	<i>“Alternative focused thinking”</i> ²⁰⁴	<i>“Value alternative hybrid thinking”</i>
Decription	Objectives and criteria selected first prior the identification or assessment of the alternative options	Options first compared so that objectives and criteria emerge based on their attributes	Generic set of objectives and criteria created first which then becomes adapted for the particular decision problem
Value tree formation	Top-down approach	Bottom-up approach	Top-down followed by bottom-up

3.3.2 Key Criteria Properties

In order for the analysis to provide the highest possible insight and to enhance its actual value to decision-makers, both criteria and attributes need to adhere to a number of key properties^{3,204-206}. If they do not, the results obtained through scoring and weighing could be spurious and, therefore, meaningless for decision-making. First, objectives or criteria need to be essential, in that all necessary objectives of the decision problem should be considered, and all the critical values under consideration should be included through the incorporation of the respective criteria. In the context of a value tree, all therapeutic, safety, burden of illness, innovation and socioeconomic criteria should be included in the model. Second, criteria need to be understandable, so that all participants in the decision-making process have a clear understanding of them and their implications. Third, criteria need to be operational; namely, the performance of the options against the criteria should be measurable. Fourth, it is crucial that criteria are non-redundant, i.e. there should be no overlap or double counting between the different criteria, otherwise the elicited criteria weights would not be accurate and, consequently the overall results would be misleading. Finally, criteria need to be concise and only the smallest set that can adequately capture the decision problem should be used, striving for simplicity and parsimony, rather than complexity.

The aggregation stage is very important because it produces the overall value scores of the alternative options. In order to enable the use of simple aggregation rules (e.g. additive value models, where scores and weights of the different individual criteria are multiplied and then added altogether in a weighted average manner), preference independence between the different criteria needs to be upheld ² (see section 1.1.4).

3.3.3 Key Attribute Properties

For selected attributes to be adequate or meaningful, sufficient properties require them to be unambiguous (in that a clear relationship should exist between the consequences of an option and the levels of attribute used to describe these consequences), comprehensive (the attribute levels should cover the full range of consequences), direct (the attribute levels should describe the consequences of alternative options as directly as possible), operational (information required for the attributes should be collectible in practice and value trade-offs - between the objectives or criteria - can be made), and understandable (the consequences and value trade-offs can be readily understood and communicated across the decision-makers and other stakeholders by using the attribute) ²⁰⁶. For an additive value model to be used, attributes should be preference independent (see section 1.1.4).

A suggested systematic methodology for selecting the best possible attributes initially involves an aim for a single natural attribute, namely one that is in general use and has a common interpretation measuring directly the degree to which an objective or a criterion is met. If no such single attribute is appropriate then a set (i.e. more than one) of natural attributes should be considered that adequately describe objective/criteria consequences. If this is not possible, exploration of ‘constructing’ attributes that directly measure consequences should be attempted. Such attributes are explicitly developed to measure directly the achievement of an objective. A proxy attribute, i.e. a less informative attribute that indirectly measures a criterion of concern, should be selected only after careful consideration and following the elimination of constructed attributes ²⁰⁶.

3.4 ‘Incremental’ Versus ‘Clean Slate’ MCDA Approaches and Link to Policy Making

The application of MCDA in current HTA practices has been criticised partly because criteria should be perceived as attributes of benefit and the fact that cost and uncertainty or quality of evidence cannot be accounted for as benefits ²⁰⁷. In turn, costs can be considered

by incorporating the ‘impact on costs’ as a criterion, other than the purchasing cost of the treatment itself, which is essentially looking at savings or increased outlays. Poor quality of evidence could be addressed through the incorporation of penalty functions that may be added when significant uncertainty exists, reducing the performance scores of relevant options. For example, if the clinical data relating to an OS gain are regarded as highly uncertain for any reasons relating to the external and/or internal validity of the clinical trial/data, then the performance score of the observed OS gain for the particular treatment could be reduced by a significant factor, e.g. 25–50 %, based on expert opinion. However, the use of penalty functions might be incompatible with the use of an additive model and therefore should be used with caution, as for example within descriptors of performance. As identified through a recent review, a number of other formal approaches also exist to quantify and incorporate uncertainty when conducting MCDA for healthcare decisions, the most commonly used being fuzzy set theory, probabilistic sensitivity analysis, deterministic sensitivity analysis, Bayesian framework and grey theory ²⁰⁸.

Another consideration relates to the appropriate process of eliciting weights, and the argument that if they are to emerge during the decision-making process it will prove difficult to achieve predictability, consistency and accountability, but also that scientific and social value judgements might become mixed and prone to strategic behaviour, with pre-specified weights being the way forward ²⁰⁷. Indeed, producing global weights that are applicable across all decision contexts would seem a very challenging task; however, weights elicited ex-ante would be hardly accurate in capturing the precise trade-offs under consideration for the reasons discussed in Section 2.2.3. By contrast, eliciting weights ex-post would be more reflective of decision-maker preferences and less susceptible to strategic manipulation; however, their application would be mostly restricted at a local decision context. A further source of criticism stems from the question of what attributes of benefit are lost due to additional cost and whether these attributes of benefit foregone can be accounted for by including the incremental cost-effectiveness ratios (ICERs) as criteria ²⁰⁷. Given that cost effectiveness and cost are not attributes of benefit, one would need to know what additional costs are required to improve the composite measure of benefit and what attributes of benefit will be given up as a consequence of costs.

Some of the above criticisms arise chiefly when MCDA is applied as a ‘supplementary’ approach to CEA to adjust the ICER by incorporating additional parameters of value. Instead, some of the above criticisms could in theory be overcome by using a ‘pure’ MCDA approach to derive value without the use of CEA (also described as

‘incremental’ and ‘clean slate’ approaches, respectively ¹⁹¹). In any case, further research would be needed to fully address some of the remaining practical limitations.

The aggregate metric of value emerging from the MCDA process (the value index) is more encompassing in nature, as multiple evaluation criteria are incorporated in the analysis. By adopting this value index metric as the benefit component and incorporating the purchasing cost of the different options, the cost per incremental MCDA value unit gained, i.e. the incremental cost value ratio(s) (ICVR), could act as the basis of allocating resources in a way comparable to that of an ICER; for instance, options with lower ICVRs would be interpreted as more valuable and could be prioritised versus options with higher ICVRs. Based on this approach, issues relating to the definition of efficiency through the establishment of thresholds that reflect opportunity costs would still need to be addressed; however, they lie outside the scope of this paper.

The resulting value index scores would be context specific, reflecting stakeholder preferences: the value index incorporates value judgements and preferences for a set of options based on a group of criteria, all of which can be informed through stakeholder input. Unless identical value judgements and preferences are assumed for the same group of criteria, a value score for an option in one setting could be different from that in another setting. The MCDA process, as proposed in this paper, respects stakeholder preferences in individual settings, whilst reducing decision-making inconsistency by introducing clarity, objectivity and greater transparency about the criteria based on which decisions can be shaped.

3.5 Conclusion

I have proposed a methodological process, with different phases and stages, outlining the use of MCDA in the context of HTA based on MAVT methods. Although a variety of MCDA techniques exist, it is likely that the most important stages that act as the foundations to the analysis are the establishment of objectives and the definition of criteria and attributes. I have focused on best practice requirements, as reflected through the appropriate properties needed for criteria and attribute selection, all of which feed into the model-building phase.

Compared with economic evaluation techniques, such as CEA, HTA through MCDA is found to have a number of important advantages. These include the multitude of criteria that can be used to assess value, the explicit weights that are assigned to reflect differences in the relative importance of the criteria, the extensive stakeholder engagement across all

stages and the transparent nature of the MCDA process, leading to a flexible and encompassing approach to value assessment and appraisal.

Chapter 4 – Paper 2

Evaluation Criteria and Decision-Making Practices in Health Technology Assessment: a Systematic Review and Expert Consultation across 8 EU countries⁷

⁷ Chapter 4 has been published with co-authors Dr Ansgar Lange and Prof Panos Kanavos as: Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *Eur J Health Econ* (2017). In press; doi:10.1007/s10198-017-0871-0

Summary

Although health technology assessment (HTA) systems base their decision-making process either on economic evaluations or comparative clinical benefit assessment, a central aim of recent approaches to value measurement, including value based assessment and pricing, points towards the incorporation of supplementary evidence and criteria that capture additional dimensions of value.

The aim of this chapter is to study the practices, processes and policies of value-assessment for new medicines across eight European countries and the role of HTA beyond economic evaluation and clinical benefit assessment. A systematic (peer review and grey) literature review was conducted using an analytical framework examining: (a) ‘Responsibilities and structure of HTA agencies’; (b) ‘Evidence and evaluation criteria considered in HTAs’; (c) ‘Methods and techniques applied in HTAs’; and (d) ‘Outcomes and implementation of HTAs’. Study countries were France, Germany, England, Sweden, Italy, Netherlands, Poland and Spain, so that the sample is representative of different health systems and HTA approaches across Europe. Evidence from the literature was validated and updated through two rounds of feedback involving primary data collection from national experts.

All countries assess similar types of evidence, however the specific criteria/endpoints used, their level of provision and requirement and the way they are incorporated (e.g. explicitly vs. implicitly) varies across countries, with their relative importance remaining generally unknown. Incorporation of additional ‘social value judgements’ (beyond clinical benefit assessment) and economic evaluation could help explain heterogeneity in coverage recommendations and inconsistency in decision-making.

More comprehensive and systematic assessment procedures characterised by increased transparency in terms of selection of evaluation criteria, their importance and their intensity of use could lead to more rational evidence-based decision-making, possibly resulting in the improvement of efficiency in resource allocation, while also raising public confidence and fairness.

4.1 Background

Current value assessment and appraisal approaches of medical technologies using economic evaluation or adopting comparative clinical benefit assessment in order to inform coverage

decisions and improve efficiency in resource allocation have been subject to criticism for a number of reasons.

Most health technology assessment (HTA) systems base their decision-making process on cost per outcome metrics of economic evaluations as, for example, the cost per QALY ¹⁸⁰. However a key limitation of the QALY approach is the inadequacy of capturing social value that goes beyond clinical outcome and cost ²⁰⁹⁻²¹¹. It is clear that a central aim of more recent approaches to value measurement, including value based assessment and value based pricing, may need to incorporate other parameters that capture different or additional dimensions of value into the overall valuation scheme ^{145,212}. Although in different HTA settings a number of additional criteria beyond scientific value judgements are incorporated to assess the evidence submitted and inform coverage decisions ²¹³, their use remains implicit or *ad hoc* rather than explicit.

Another drawback is caused through the way which value is assessed and appraised, often resulting in unexplained heterogeneity of coverage decisions across settings even for the same drug-indication pair ^{158-160,214-217}. Although some of this decision-making differentiation can be justified on the grounds of different budget constraints and national priorities, inconsistencies of medicines' eligibility for reimbursement across countries give rise to an international 'post-code' lottery for medicines access even in the same geographical region, with important implications for equity and fairness especially when differences remain unexplained ²¹⁵. Several studies have acknowledged the need for well-defined decision-making processes that are fairer and more explicit ²¹⁸⁻²²⁰. By ensuring 'accountability for reasonableness' and providing a better understanding of the rationale behind decision-making, decisions will also have enhanced legitimacy and acceptability ^{19,216}.

By reviewing and synthesising the evidentiary requirements (both explicit and implicit), the methods and techniques applied and how they lead to decisions, the objective of this study is to provide a critical review of value assessment and appraisal methods for new medicines including the evaluation criteria employed, across a number of jurisdictions in Europe that employ explicit evaluation frameworks as part of their HTA processes. In more detail, the study seeks to determine whether HTA processes incorporate additional criteria beyond economic evaluation or clinical benefit assessment and, if so, which ones and how they inform coverage recommendations. To my knowledge, to date no study has provided a similar review and analysis of HTA policies and practices for innovative medicines across different European countries to this extent. In fulfilling the above aims, the

next section outlines the methods and includes the components of the analytical framework adopted for this purpose; subsequently, the evidence collected from eight European countries is presented and discussed, before presenting the policy implications.

4.2 Methods

An analytical framework is adopted to facilitate the systematic review of HTA processes and capture their salient features across settings following previous evidence²²¹. Based on that, the relevant evidence was collected relying both on primary and secondary sources. The evidence base covered eight EU Member States that have arms-length HTA bodies and recognised HTA processes. The study took place in the context of Advance-HTA, an EU-funded project aiming to contribute to advances in the methods and practices for HTA in European and elsewhere²²².

Secondary sources of evidence comprised a systematic review of the country-specific value-assessment peer review literature using the newly adapted analytical framework to investigate the practices, processes and policies of value-assessment and their impact, as observed in the study countries.

Evidence from the literature was validated by means of two rounds of feedback involving primary data collection: the first was from Advance-HTA consortium partners²²², while the second involved a detailed validation of the study's results by national experts following the incorporation of all literature results and feedback from Advance-HTA partners.

4.2.1 Analytical Framework Outlining the Components of a Value Assessment Framework

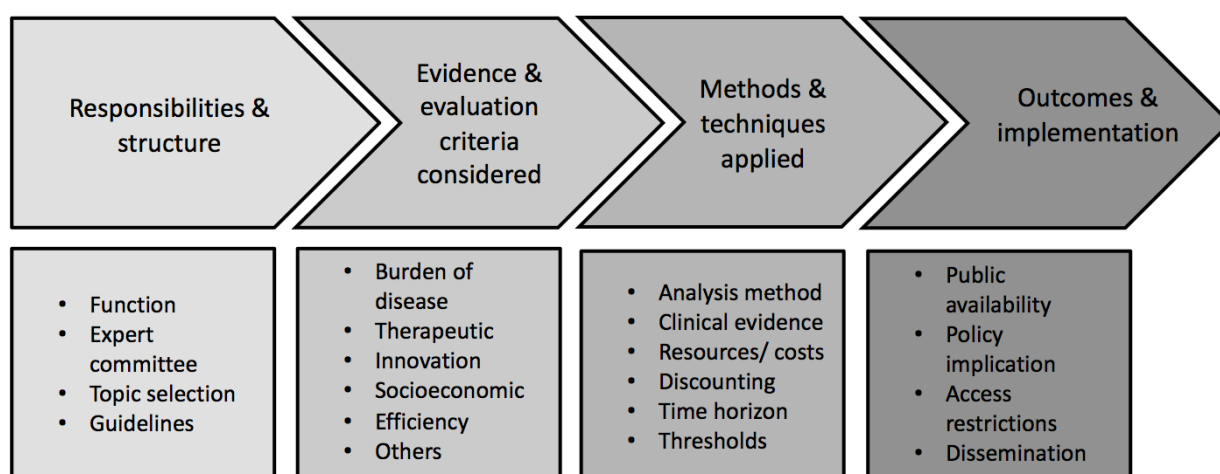
Existing frameworks for analysing and classifying coverage decision-making systems for health technologies were reviewed and adjusted according to the needs of the current examination which focuses on the assessment and appraisal stages of the coverage review procedure from the HTA institution's point of view, without having a special interest on the decision outcomes *per se*, or their translation into reimbursement decisions²²³⁻²²⁵.

The main value assessment characteristics necessary to outline the practices and processes different countries of interest as reflected through their HTA bodies were classified using four key components, each of them having a number of different (sub-) components: (a) 'Responsibilities and structure of HTA agencies'; (b) 'Evidence and evaluation criteria considered in HTAs'; (c) 'Methods and techniques applied in HTAs'; and (d) 'Outcomes

and implementation of HTAs' were considered to be the main components needed in order to sufficiently capture the features of the different value assessment systems.

In the context of the study the second component, (b), carried greater weight because a key subject of the investigation was to explore and analyse the different evaluation criteria beyond those informing economic evaluations or clinical benefit assessment. The sub-components of the main components are described below and are shown in Figure 4.1.

Figure 4.1: Main components and sub-components of the analytical framework applied



4.2.1.1 Responsibilities and structure of national HTA agencies

This component considers the operational characteristics of HTA agencies. It includes details about the function and responsibilities of HTA bodies, the relevant committees within agencies tasked with assessment and appraisal, details on the topic selection process, and whether methodological guidelines exist for the conduct of pharmacoeconomic analysis.

4.2.1.2 Evidence and evaluation criteria considered in HTAs

This component relates to the types of evidence evaluated and the particular evaluation criteria considered. Generally, the assessed evidence can be classified into features relating to the disease (indication) under consideration, or into characteristics relating to the technology being assessed. The former is reflected through 'Burden of Disease' (BoD), i.e. the impact that the disease has, which mainly depends on the severity of the disease and the unmet medical need. The latter group of a technology's characteristics can be classified into clinical benefit (mainly therapeutic impact and safety considerations), innovation (e.g.

clinical novelty and nature of treatment), and socioeconomic impact (e.g. public health impact, productivity loss impact) categories. Other characteristics relating to efficiency (e.g. cost-effectiveness, cost), ethical/equity considerations, accepted data sources, and relative importance (i.e. weighting) of the evidence are also listed.

4.2.1.3 Methods and techniques applied in HTAs

This component is associated with the evaluation methods and techniques used. In terms of the analytical methods applied (i.e. comparative efficacy/effectiveness, type of economic evaluation), methodologies differ according to their outcome measure and their elicitation technique, the choice of the comparator(s), and the perspective adopted. In relation to the clinical evidence used to populate the analysis, crucial details involve preferred data sources (i.e. study designs), data collection approaches (e.g. requirement for systematic literature reviews) and synthesis (e.g. suggestion for meta-analysis) of the data. In terms of resources used, important specifics include the types of costs included and data sources used. For both clinical outcomes and costs, discount rate(s) applied and time horizons assumed are included, together with the existence of any explicit or implicit cost-effectiveness thresholds based on which decisions are made.

4.2.1.4 Outcomes and implementation of HTAs

The final component relates to the outcomes of the evaluation procedures and their implementation. Key characteristics include the public availability of the evaluation report, the policy implications and whether outcomes are applied in practice (e.g. pricing vs. reimbursement), the usage of any access restrictions, how are decisions disseminated and implemented, whether appeal procedures are available and the frequency of decision revisions.

4.2.2 Systematic Literature Review

The systematic literature review methodology was based on the Centre for Review and Dissemination (CRD) guidance for undertaking systematic reviews in health care ²²⁶.

4.2.2.1 Inclusion criteria (country selection and study period)

The study countries (and the respective agencies) were France (Haute Autorité de Santé, HAS), Germany (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, IQWiG), Sweden (Tandvårds- och läkemedelsförmånsverket, TLV), England (National Institute for Health and Care Excellence, NICE), Italy⁸ (Agenzia Italiana del Farmaco, AIFA), Netherlands (Zorginstituut Nederland, ZIN (formerly College voor zorgverzekeringen, CVZ)), Poland (The Agency for Health Technology Assessment and Tariff System, AOTMiT) and Spain (Red de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) and the Interministerial Committee for Pricing (ICP))⁹. The study countries were selected because of the variation in their health system financing (tax-based vs. social insurance-based), the organisation of the health care system (central vs. regional organisation), the type of HTA in place (predominantly economic evaluation vs. predominantly clinical benefit assessment) and the perspective used in HTA (health system vs societal), so that the sample is representative of different health systems and HTA approaches across Europe.

The study period for inclusion of relevant published studies was from January 2000 to January 2014, with article searches taking place in February 2013 in the first instance and an update taking place at the end of January 2014. The year of 2000 was selected as the start date because the HTA activity of most countries started or significantly expanded in scope since then. Feedback from the Advance-HTA consortium partners was provided in August 2014. Additional input, including the most recent updates on national HTA processes, was also collected from HTA experts and national competent authorities.

4.2.2.2 Identification of evidence

The electronic databases of MEDLINE (through PubMed resource) and Social Science Citation Index (through Web of Science portal) were searched for peer review literature only using a search strategy for English articles published up until the time of the literature search (including all results from the oldest to the latest available) using the following keywords:

⁸ Other HTA bodies exist on regional level (e.g. UVEF is responsible for HTAs in the Veneto region).

⁹ RedETS is the Spanish Network of regional HTA agencies coordinated by Instituto de Salud Carlos III (ISCIII) and could be regarded as the National HTA advisory body at central level. However at national level it does not assess pharmaceuticals but mostly non-drug health technologies such as screening programmes and medical devices. Instead, the Interministerial Committee for Pricing (ICP), led by the Dirección General de Farmacia under the Ministry of Health, is the committee responsible for the assessment of drugs, producing mandatory decisions at national level regarding the reimbursement and pricing of pharmaceuticals.

‘health technology assessment + pharmaceuticals’; ‘health technology assessment + methodologies’; ‘value assessment + pharmaceuticals’; ‘value assessment + methodologies’. In addition, reference lists from the studies selected were screened (see below, ‘study selection & data extraction’), retrieving any additional studies that could be potentially relevant. Furthermore, grey literature was searched including published guidelines from the HTA bodies available online through each agency’s website.

4.2.2.3 Study selection & data extraction

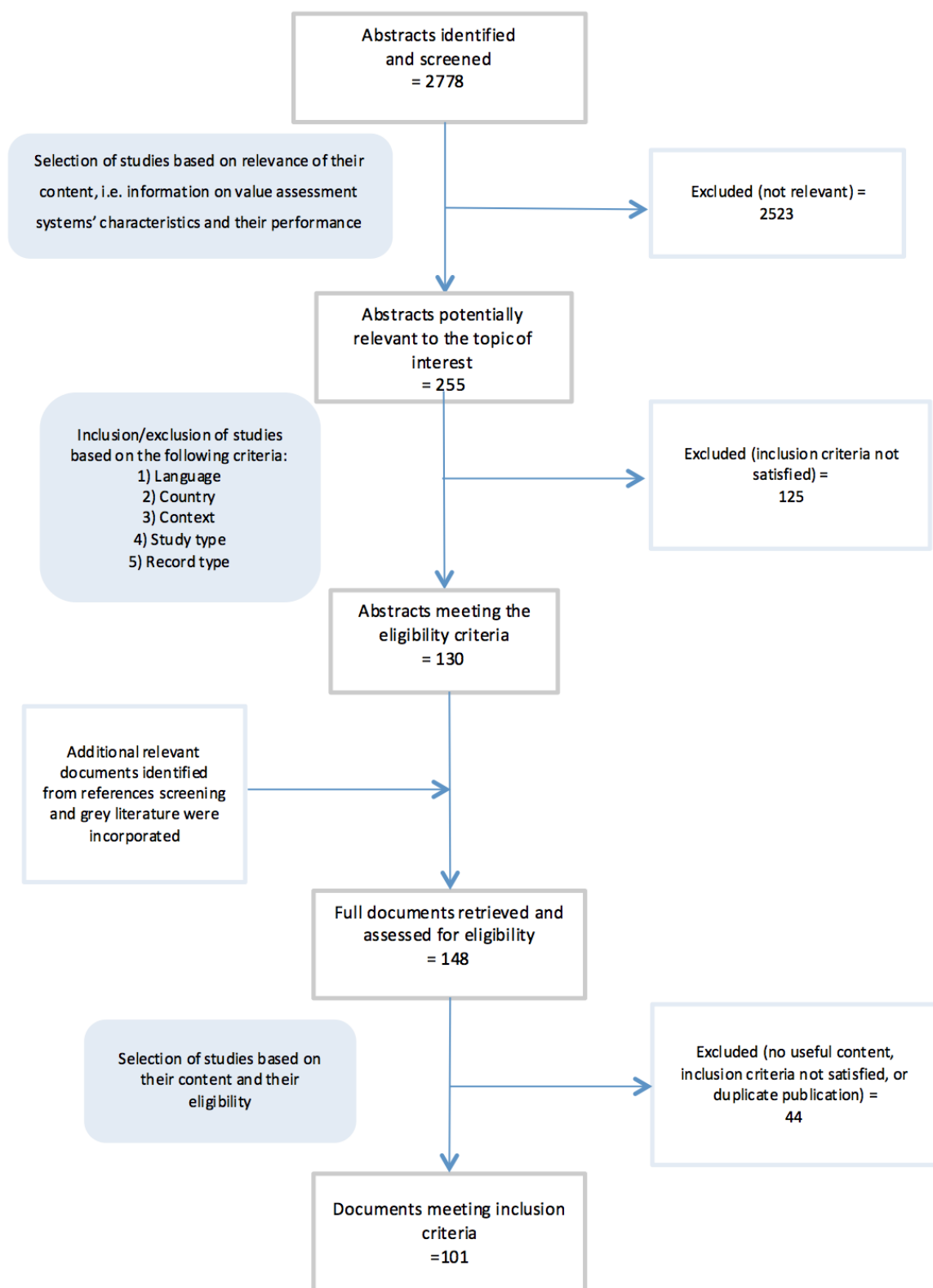
Articles were selected according to a four-stage process as outlined in Figure 4.2 ²²⁶. As part of the first stage, all titles and abstracts were reviewed, with abstracts not relevant to the topic excluded; in cases where content relevance could not be determined articles were passed through the next stage. In the second stage, all relevant abstracts were assessed against a number of pre-determined selection criteria; these included: i) language (only English articles included), ii) study country (only studies examining the eight countries of interest included), iii) study context (only national coverage HTA perspectives included), iv) study type (product-specific technology appraisal reports were excluded), v) record type (conference proceedings or titles with no abstracts available were excluded). In the third stage, full articles for all abstracts meeting the eligibility criteria were retrieved; in addition, relevant studies identified from reference screening and grey literature, including published guidelines from HTA agencies, were incorporated (non-English articles cited by English documents were included at this stage). Finally, in the fourth stage full articles were reviewed and relevant data were extracted. An Excel template listing the value assessment components (categories and sub-categories) of interest was used for data extraction. Data were extracted in free text form, with no limitations on the number of free text fields and as little categorisation of data as possible, in order to avoid loss of information.

4.2.3 Expert Consultation

Upon consultation of the preliminary results with the partners of the Advance-HTA project consortium, it became obvious that in a few cases (primarily in France and, to a lesser degree in Sweden), the evidence from the peer review literature did not reflect actual practices being even contradictory in some cases. As a result, comments and feedback were solicited from the project’s consortium partners. Finally, updated results tables were subsequently shared with HTA experts in the study countries which were asked to review and validate the outputs of the study. Experts were affiliated with academic research institutions or national

competent authorities and provided further evidence and guidance, including additional literature sources outside the originally selected review period, if applicable. Experts' inputs from these two rounds of consultation are quoted as personal communications from the Advance-HTA project ²²⁷.

Figure 4.2: Flow chart of the systematic literature review process



4.3 Results

Figure 4.2 shows a flow chart of the review process and the resulting number of articles in each stage. In total, 2778 potentially eligible peer review articles were identified in the electronic databases and 255 articles were identified as potentially useful and were read in full. Of these, 130 articles met the eligibility criteria and an additional 18 articles were identified as possibly relevant through reference screening and grey literature. The content of 101 articles from the literature review was finally used to inform the findings ^{65,76,124,131,216,224,228-321}. An additional five studies were identified during the expert consultation process and were taken into consideration in discussing and interpreting the results ³²²⁻³²⁶.

4.3.1 Responsibilities and Structure of National HTA Agencies

Across the study countries, HTA agencies mainly exist in the form of autonomous governmental bodies, having either an advisory or regulatory function. Usually, a technical group is responsible for early assessment of the evidence following which an expert committee appraises the request for coverage and produces recommendation(s) for the final decision body.

The topic selection process is generally not entirely-transparent, with the belief that most agencies predominantly assess new medical technologies, and more precisely medicines, that are expensive and/or with uncertain benefits. In some cases, topic selection is not applicable as all technologies that apply for reimbursement need to be assessed.

In all study countries, official country-specific pharmacoeconomic guidelines for the evaluation process are available, mainly concerning methodological and reporting issues ^{267,284}; in England, in addition to the evaluation process, guidelines also exist for the purpose of application submission requirements, including the description of key principles of the appraisal methodology adopted by NICE ²⁶⁷. For all countries the application of the guidelines is recommended. A summary of the responsibilities and structure of the national HTA agencies in the study countries is presented in Table 4.1.

4.3.2 Evidence and Evaluation Criteria Considered in HTAs

Generally all countries assess the same groups of evidence, however the individual parameters considered and the ways they are evaluated differ from country to country. All countries acknowledge the existence of a wide variety of data sources including scientific

studies (e.g. clinical trials, observational studies), national statistics, clinical practice guidelines, registry data, surveys, expert opinion and evidence from pharmaceutical manufacturers²⁸³. A summary of the evidence and the evaluation criteria under consideration across the study countries is presented in Table 4.2.

4.3.2.1 Evaluation principles and their relevance to priority setting

In France the assessment of the product's medical benefit (Service Médical Rendu, SMR) and improvement of medical benefit (Amélioration du Service Médical Rendu, ASMR) determine the reimbursement and pricing of the drug respectively. As of October 2013, economic criteria have been introduced with the Commission for Economic Evaluation and Public Health (CEESP) evaluating the cost-effectiveness (without a cost-effectiveness threshold in place) of products assessed to have an ASMR I, II or III that are likely to impact social health insurance expenditures significantly (total budget impact greater than EUR 20 million), being used by the Economic Committee for Health Products (CEPS) in its price negotiations with manufacturers³²⁷. Nevertheless, and under this current framework, these economic evaluations have not the same impact on price negotiation with ASMR which are directly linked with pricing but instead their role is limited to a consultative one.

Table 4.1: Responsibilities and structure of national HTA agencies

	FRANCE (HAS / CEESP)	GERMANY (IQWiG)	SWEDEN (TLV)	ENGLAND (NICE)	ITALY (AIFA)	NETHERLANDS (ZIN)	POLAND (AOTMiT)	SPAIN (RedETS/ISCIII or ICP¹⁰)
Function	Autonomous Advisory	Autonomous, Advisory	Autonomous, Regulatory	Autonomous, Advisory	Autonomous, Regulatory	Autonomous, Advisory	Autonomous, Advisory	Autonomous, Advisory
Expert committee	Transparency Commission, Economic Evaluation and Public Health Commission (CEESP)	Assessment: IQWiG scientific personnel; Appraisal: Federal Joint Committee (G-BA)	The Board for Pharmaceutical Benefits	Technology Appraisal Committee	AIFA's Technical Scientific Committee and AIFA's Pricing and Reimbursement Committee (CPR)	Committee for societal consultation regarding the benefit basket	Transparency Council	Interministerial Committee for Pricing ¹¹ (ICP)
Topic selection	HAS (About 90% submitted by the manufacturers, 10% requested by the MoH) ¹²	Not applicable (all drugs applying for marketing authorization as of 2011)	TLV (only outpatient and high price drugs)	DH in consultation with NICE based on explicit prioritisation criteria ¹³	AIFA (all drugs submitted by manufacturers)	Mostly on its own initiative; sometimes at the request of MoH	MoH ¹⁴ (in the case of manufacturer submission – triggered by MAH)	Not subject to any specific known procedure ¹⁵
Guidelines for economic analysis	Yes	Yes (however, CBA is not standard practice)	Yes	Yes	In progress	Yes	Yes	Spanish recommendations on Economic Evaluation of Health Technologies

¹⁰ RedETS is the Spanish Network of regional HTA agencies, coordinated by ISCIII, responsible for the evaluation of non-drug health technologies. The ICP, led by the Dirección General de Farmacia under the Ministry of Health, is the committee responsible for the evaluation of drugs producing mandatory decisions at national level.

¹¹ The ICP involves representatives from the Ministry of Health, Ministry of Industry and Ministry of Finance together with a dynamic (i.e. rotating) set of expert representatives from the autonomous communities.

¹² An economic evaluation is performed only for a subset of new products meeting certain criteria (manufacturer claims a high added value / product is likely to have a significant impact on public health expenditures).

¹³ Criteria include expected health benefit, population size, disease severity, resource impact, inappropriate variation in use and expected value of conducting a NICE technology appraisal.

¹⁴ Regulated by law: the Act of August 27th 2004 on healthcare benefits financed from public funds; the Act of May 12th 2011 on the reimbursement of medicinal products, special purpose dietary supplements and medical devices.

¹⁵ For new drugs, manufacturers have to submit a dossier for evaluation when they apply for pricing and reimbursement. Topic selection for non-drug technologies under the action of RedETS is well developed with the participation of informants from all autonomous communities based on a two round consultation

Table 4.2: Evidence and evaluation criteria considered in HTAs

Burden of Disease								
	FRANCE (HAS / CEESP)	GERMANY (IQWiG)	SWEDEN (TLV)	ENGLAND (NICE)	ITALY (AIFA)	NETHERLANDS (ZIN)	POLAND (AOTMiT)	SPAIN (RedETS/ISCIII or ICP)
Severity	Yes, as part of SMR	Yes, as part of added benefit assessment	Yes (impact on WTP threshold) ¹⁶	Yes (mainly as part of EoL treatments)	Yes (implicitly)	Yes ¹⁷	Yes ¹⁸	Yes
Availability of treatments (i.e. unmet need)	Yes (binary: Yes/No)	True for other technologies rather than pharmaceuticals ¹⁹	Yes, indirectly (captured by severity)	Yes (clinical need as a formal criterion)	Yes ²⁰	Yes ²¹	Yes ²²	Yes
Prevalence (e.g. rarity)	Yes, informally	As part of G-BA's ²³ decision-making process	Yes	Yes	Yes ²⁴	Yes	Yes ²⁵	Yes
Therapeutic & Safety impact								

¹⁶ Severity can be defined on the basis of several elements of the condition, including the risk of permanent injury and death.

¹⁷ Both explicitly and implicitly; more recently they tend to explicitly take into account “burden of disease” measures.

¹⁸ Regulated by law: the Act of August 27th 2004 on healthcare benefits financed from public funds.

¹⁹ In evaluations performed by the G-BA to determine the benefit basket (i.e. not drugs, which are covered automatically after marketing authorization and value assessment plays a role for the price) availability or lack of alternatives and the resulting medical necessity are considered to determine clinical benefit.

²⁰ Explicitly stated in the legislation as a criterion to set price.

²¹ Estimate the number of treatments that is considered necessary and compared that with the actual capacity.

²² Not obligatory by law; considered in the assessment process in AOTMiT on the base of HTA Guidelines (good HTA practices).

²³ Lower accepted significance levels for p-values (e.g. 10% significance levels) for small sample sizes such as rare disease populations; acceptance of evidence from surrogate endpoints rather than only ‘hard’ endpoints; additional benefit is considered proven at marketing authorisation if budget impact is less than €50m per annum.

²⁴ Decisions on price and reimbursement of orphan drugs are made through a 100-day *ad hoc* accelerated procedure, although criteria for HTA appraisals do not differ from non-orphan drugs.

²⁵ Commonness, but not rarity, regulated by law (the Act on healthcare benefits); rarity is considered in the assessment process in AOTMiT on the base of HTA Guidelines.

Efficacy	Yes (4 classifications via SMR, 5 via ASMR) ²⁶	Yes (6 classifications) ²⁷	Yes	Yes	Yes	Yes	Yes ²⁸	Yes
Clinically meaningful outcomes	Yes (preferred)	Yes (preferred)	Yes	Yes (preferred)	Yes	Yes	Yes ²⁹	Yes
Surrogate/intermediate outcomes	Considered	Considered	Considered	Considered	Considered	Considered	Considered ³⁰	Considered
HRQoL outcomes	Generic; disease-specific	Not specified ³¹	Generic (preferred); disease-specific	Generic; disease-specific	Generic; disease-specific	Yes	Yes ³²	Yes (including patient well-being)
Safety	Yes	Yes ³³	Yes	Yes	Yes	Yes	Yes ³⁴	Yes
Dealing with uncertainty	Implicitly (preference for RCTs), Explicitly	Explicitly (classification of empirical studies and complete evidence)	Implicitly (through preference for RCTs)	Explicitly (quality of evidence), Implicitly (preference for RCTs), Indirectly	Yes, registries and MEAs are used to address uncertainty	Implicitly (if included in the assessment studies)	No ³⁵	Can be considered as part of economic evaluations

²⁶ SMR, 4 classifications for actual clinical benefit: Important/High (65% reimbursement rate), Moderate (30%), Mild/Low (15%), Insufficient (not included on the positive list); ASMR, 5 classifications for relative added clinical value: Major (ASMR I), Important (ASMR II), Moderate (ASMR III), Minor (ASMR IV), No clinical improvement (ASMR V).

²⁷ The possible categories are: major added benefit, considerable added benefit and minor added benefit. Three additional categories are recognized: non-quantifiable added benefit, no added benefit, and lesser benefit.

²⁸ Regulated by law: the Act of August 27th 2004 on healthcare benefits financed from public funds.

²⁹ Regulated by law: the Act on the reimbursement.

³⁰ Weak preference; if no LYG/QALY data available.

³¹ Considered if measured using validated instruments employed in the context of clinical trials.

³² Regulated by law: the Act on the reimbursement.

³³ Based on the following ranking relative to comparator: greater harm, comparable harm, lesser harm

³⁴ Regulated by law: the Act on healthcare benefits; the Act on the reimbursement.

³⁵ Not obligatory by law; considered in the assessment process in AOTMiT on the base of HTA Guidelines (good HTA practices).

	(robustness of evidence)			(rejection if not scientifically robust)				
Innovation level								
Clinical novelty	Yes (as part of ASMR) if efficacy/ safety ratio is positive	Implicitly as part of added therapeutic benefit consideration ³⁶	Yes, but only if it can be captured in the CE analysis	Yes	Yes	Yes	Yes ³⁷	Yes ³⁸
Ease of use and comfort	Not explicitly, in some cases ³⁹	Not explicitly considered for benefit assessment ⁴⁰	Yes (to some extent)	Not explicitly	No	Not standard, case-by-case basis	No ⁴¹	Not explicitly, indirectly ⁴²
Nature of treatment/technology	Yes (3 classifications) ⁴³	Not explicitly considered for benefit assessment	Not explicitly	Yes (when above £20,000)	No	Implicitly	Yes ⁴⁴	Yes (through the degree of innovation criterion)
Socio-economic impact								
Public Health Benefit/Value	Yes, rarely via "intérêt de	Yes ⁴⁶	Yes, indirectly ⁴⁷	As indicated in guidance to NICE	Implicitly	Yes (explicit estimates)	Yes ⁴⁹	Social utility of the drug and

³⁶ Not a criterion *per se*, implicitly considered if patient benefit is higher than that of existing alternatives.

³⁷ The Act on healthcare benefits considers the following classifications: saving life and curative, saving life and improving outcomes, preventing the premature death, improving HRQoL without life prolongation.

³⁸ Incremental clinical benefit is considered as part of the therapeutic and social usefulness criterion.

³⁹ Only considered in the ASMR if it has a clinical impact (e.g. through a better compliance).

⁴⁰ The IQWiG's general methodology (not specifically for new drugs) states that patient satisfaction can be considered as an additional aspect, but it is not adequate as a sole deciding factor.

⁴¹ Not obligatory by law (unless captured in HRQoL/QALY); considered in the assessment process in AOTMiT on the base of HTA Guidelines.

⁴² Through the therapeutic and social usefulness criterion.

⁴³ Ranking includes the following classifications: Symptomatic relief, Preventive treatment, Curative therapy.

⁴⁴ Regulated by law: the Act on healthcare benefits considering the following classifications: saving life and curative, saving life and improving outcomes, preventing the premature death, improving HRQoL without life prolongation; thus no "innovativeness" *per se*.

⁴⁶ Manufacturer dossiers need to include information on the expected number of patients and patient groups for which an added benefit exists as well as costs for the public health system (statutory health insurance).

⁴⁷ The following principles are considered: Human dignity, Need/solidarity, Cost-efficiency, Societal view.

⁴⁹ Regulated by law: the Act on healthcare benefits considering: impact on public health in terms of priorities for public health set; impact on prevalence, incidence – qualitative assessment rather than quantitative.

	Santé Publique ⁴⁵			to be considered in the evaluation process ⁴⁸				rationalisation of public drug expenditures
Social Productivity	Not explicitly ⁵⁰	Yes ⁵¹	Indirect costs considered explicitly (to some extent)	Considered, but not explicitly incorporated as part of ICERs	Direct costs only ⁵²	Yes	No ⁵³	Yes, either explicitly or implicitly
Efficiency considerations								
Cost-Effectiveness	Yes ⁵⁴	Yes (cost-benefit) ⁵⁵	Yes (cost-efficiency as a principle)	Yes	Yes	Yes	Yes, mandatory by law	Yes (not mandatory)
Cost/Budget Impact Analysis	Not mandatory but BIA is highly recommended ⁵⁶	BIA (mandatory)	Cost only considered for treatments of the same condition; BIA not mandatory	BI to NHS, PSS, hospitals, primary care	Yes	Yes	Yes, payer affordability mandatory by law	Yes (BI to NHS)
Other evidence and criteria								

⁴⁵ Public health interest (intérêt santé publique; ISP) is incorporated into the SMR evaluation. ISP considers 3 things: whether the drug contributes to a notable improvement in population health; whether it responds to an identified public health need (e.g. ministerial plans); and whether it allows resources to be reallocated to improve population health.

⁴⁸ Factors include cost-effectiveness, clinical need, broad priorities for NHS, effective use of resources and encouragement of innovation, and any other guidance issued by the Secretary of State.

⁵⁰ Only potentially as part of economic evaluations.

⁵¹ Productivity loss due to incapacity as part of the cost side, productivity loss due to mortality as part of the benefit side (no unpaid work, e.g. housework).

⁵² Indirect costs can be taken into account in a separate analysis.

⁵³ No social perspective obligatory by law; may be provided but problematic to use for recommendation/decision.

⁵⁴ Already implemented but analysis conducted separately by the distinct Commission for Economic Evaluation and Public Health (CEESP). The health economic evaluation does not impact the reimbursement decision.

⁵⁵ Cost-benefit analysis is not standard practice in the evaluation but rather can be initiated if no agreement is reached between sickness funds and manufacturer on the price premium or if the manufacturer does not agree with the decision of the G-BA regarding premium pricing (added benefit).

⁵⁶ ASMR V drugs should be listed only if they reduce costs (lower price than comparators or induce cost savings).

Place in therapeutic strategy	Yes ⁵⁷	Evaluation usually specifies the line of treatment	Evaluation usually specifies the line of treatment	Broad clinical priorities for the NHS (by Secretary of State)	Yes	Not explicitly	No	Yes ⁵⁸
Conditions of use	Yes (e.g. the medicine is assessed in each of its indications, if several)	No, drug is in principle reimbursable for the whole indication spectrum listed on its authorisation ⁵⁹	Yes, coverage can be restricted based on evidence at sub-population level	Yes, coverage can be restricted based on evidence at sub-population level	Implicitly	Yes, indications	Yes, coverage can be restricted to strictly defined sub-populations	Yes (several medicines are introduced with Visado – Prior Authorization Status)
Ethical considerations	Not incorporated in assessment ⁶⁰	Sometimes (implicitly)	Yes	Yes ⁶¹	Implicitly	Yes, explicitly (e.g. solidarity and affordability) 62	Considered on the basis of HTA Guidelines	Not explicitly
Weights of different criteria	Not transparent	Not transparent	“Human dignity” usually being overriding ⁶³	Not transparent	Not transparent	Therapeutic value is the most important criterion	Not transparent	Not transparent and not consistent across regions ⁶⁴

⁵⁷ The commission will also make a statement on if a drug shall be used as first choice or only if other existing therapeutics are not effective in a patient.

⁵⁸ In the form of the new IPT – Informes de Posicionamiento Terapéutico /Therapeutic Positioning report.

⁵⁹ Sub-groups are examined during as part of benefit assessment but in order to guide pricing, not reimbursement eligibility. If a drug has an added benefit for some groups but not for others, a so-called “mixed price” is set that reflects both its added benefit for some patients and lack thereof for others.

⁶⁰ The assessment in France is purely ‘scientific’ i.e. focuses on the absolute and comparative merits of the new therapy and its placement in the therapeutic strategy.

⁶¹ NICE principles include fair distribution of health resources, actively targeting inequalities (SoVJ); equality, non-discrimination and autonomy.

⁶² Also indirectly through an ethicist seat in the Committee.

⁶³ No clear order between “need & solidarity” and cost-efficiency. In the entire health system a more complete ordering is seen where human dignity takes precedence over the principles of need & solidarity, which takes precedence over cost-efficiency.

⁶⁴ Not all regions have neither HTA bodies nor regional committees for drug assessment. However, at regional level drugs assessment is limited to prioritizing (or not) its use by means of guidelines or protocols together with some type of incentives to promote savings.

Accepted data sources (for estimating number of patients, clinical benefits and costs)	Clinical trials, observational studies, national statistics, clinical guidelines, surveys, expert opinions	RCTs ⁶⁵ , national or local statistics, clinical guidelines, surveys, price lists, expert opinions	Clinical trials, observational studies, national statistics, clinical guidelines, surveys, expert opinions	Clinical trials, observational studies, national or local statistics, clinical guidelines, surveys, expert opinions	Clinical trials, observational studies, national statistics, clinical guidelines, surveys, expert opinions, scientific societies' opinion	Clinical trials, clinical guidelines, expert opinions	Clinical trials, observational studies, national or local statistics, clinical guidelines, surveys, expert opinions	Clinical trials, observational studies, national statistics, clinical guidelines, surveys, expert opinions
--	--	---	--	---	---	---	---	--

⁶⁵ For therapeutic benefit, other designs such as non-randomised or observational studies might be accepted in exceptional cases if properly justified, e.g. in the case that RCTs are not possible to be conducted, if there is a strong preference for a specific therapeutic alternative on behalf of doctors or patients, if other study designs can provide sufficiently robust data, etc.

In Germany the new Act to Reorganize the Pharmaceuticals Market in the Statutory Health Insurance (SHI) System [Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (AMNOG)] came into effect on 1 January 2011. Since then, all newly introduced drugs are subject to early benefit assessment. Pharmaceutical manufacturers have to submit a benefit dossier for evaluation by the Institute for Quality and Efficiency in Health Care (IQWiG). A final decision is made by the Federal Joint Committee (Gemeinsame Bundesausschuss, G-BA). Benefit for new drugs encompasses the “patient-relevant therapeutic effect, specifically regarding the amelioration of health status, the reduction of disease duration, the extension of survival, the decrease in side effects or the improvement of quality of life”³²³. Importantly, all new drugs are reimbursed upon marketing authorisation and benefit assessment mainly determines price rather than reimbursement status.

In Sweden, a prioritisation framework with three explicit factors for the allocation of resources is used: i) human dignity; ii) need and solidarity; and iii) cost-efficiency^{238,249,271,294}. However, in the specific legislation for the pharmaceutical reimbursement system, human value is generally seen as the overriding criterion with no clear order between the other two²²⁷. Marginal benefit or utility, according to which a diminishing cost-effectiveness across indications and patient groups is explicitly recognized, could be regarded as a fourth principle mainly meaning that there are no alternative treatments that are significantly more suitable^{294,312,313}.

In England, the Secretary of State for Health has indicated to NICE a number of factors that should be considered in the evaluation process: (i) the broad balance between benefits and costs (i.e. cost-effectiveness); (ii) the degree of clinical need of patients; (iii) the broad clinical priorities for the NHS; (iv) the effective use of resources and the encouragement of innovation; and (v) any guidance issued by the Secretary of State^{237,268,287}. Decisions are supposed to reflect society’s values, underlined by a fundamental social value judgment²⁹⁶.

The Netherlands focuses on four priority principles when assessing medical technologies: (a) the “necessity” of a drug (severity / burden of disease)^{65,242}; (b) the “effectiveness” of a drug, according to the principles of Evidence-Based Medicine (EBM)^{242,253}; (c) the “cost-effectiveness” of a drug²³¹; and (d) “feasibility”, that is how feasible and sustainable is to include the intervention or care provision in the benefit package^{232,248}.

In Italy, reimbursement of pharmaceuticals at the central level is evaluated by AIFA's Pricing and Reimbursement Committee (CPR) which sets prices and reimbursement conditions for drugs with a marketing authorisation based on evidence of the following factors: product's therapeutic value (cost/efficacy analysis) and safety (pharmacovigilance), degree of therapeutic innovation, internal market forecasts (number of potential patients and expected sales), price of similar products within the same or similar therapeutic category, product prices in other European Union Member States ²²⁷. In autonomous regions, pricing and reimbursement of new drugs does not require –except for very innovative drugs– epidemiologic or economic evaluation studies nor assessment of cost impact from adoption of new drugs, as in other countries ^{227,280}.

An HTA in Poland is considered complete if it contains (a) a clinical effectiveness analysis; (b) an economic analysis; and (c) a healthcare system impact analysis. No studies derived from the search referred to the evidence assessed or the different parameters considered by AHTAPol (now AOTMiT) in Poland in more detail ²⁹⁰.

Finally, in Spain, different regions apply a range of different assessment requirements, but in general four main evidence parameters are considered: (a) the severity of the disease; (b) the therapeutic value and efficacy of the product; (c) the price of the product; and (d) the budget impact for the Spanish National Health System. The assessment is usually a classification or cost-consequences analysis that does not take into account the long-term effects of a therapy or the possible need of specialized care utilisation. Patient well-being and quality of life are also considered ²³⁰.

4.3.2.2 Evaluation criteria taken into account in HTAs

Burden of Disease

In France, both the severity and the existence of alternative treatments are acting as formal criteria, thus essentially defining the concept of 'need' ⁶⁵. Severity is considered as part of the SMR, taking into account symptoms, possible consequences (including physical or cognitive handicap) and disease progression (in terms of mortality and morbidity) ²²⁷. The existence of alternatives is scored against a categorical 2-level scale (yes vs. no) ^{258,259}.

In Germany, severity is considered as part of added (clinical) benefit assessment. The clinical assessment is based on "patient-relevant" outcomes, mainly relating to how the patient survives, functions or feels, essentially accounting for the dimensions of mortality, morbidity and health-related quality of life (HRQoL) ²⁶⁵.

In Sweden, severity of the condition and availability of treatments (reflected through marginal benefit / utility as a sub-principle) appear to be two of the primary criteria for priority-setting with more severe indications being explicitly prioritized via greater willingness to pay ^{65,294,312,313}.

In England, the degree of unmet clinical need is a formal criterion taken into account, at least partially being reflected by the availability of alternative treatments ^{65,315}. NICE acknowledges that rarity has a key role in the assessment of orphans and NICE's Citizens' Council has stated that society would be willing to pay more for rare and serious diseases ²⁶³. The severity of the disease is taken into account mainly through the special status of life-extending medicines for patients with short-life expectancy as reflected through the issuing of supplementary advice of life extending end-of-life (EOL) treatments by NICE ^{270,315}.

Severity of disease, availability of treatments and prevalence of the disease is generally considered across the remaining countries, either explicitly or implicitly, although not always as mandatory requirements by law but just good HTA practices (e.g. Poland for the case of treatments availability) ²²⁷.

Therapeutic impact and safety

Clinical evidence relating to therapeutic efficacy and safety acts as the most important formal criteria of the evaluation process in France ³²⁸. The product's medical benefit or medical service rendered (SMR) relates to the actual clinical benefit of the drug, responding to the question of whether the drug is of sufficient interest to be covered by social health insurance. It takes into consideration the following criteria: (a) the seriousness of the condition; (b) the efficacy of the treatment; (c) side effects of the drug; (d) its place within the therapeutic strategy given other available therapies; and (e) its interest for public health ^{227,267}.

Similarly to France, in Germany all clinically relevant outcomes are considered and final clinically meaningful outcomes (e.g. increase in overall survival, reduction of disease duration, improvement in HRQoL) are preferred over surrogate and composite endpoints ^{265,267,273,283,302}. HRQoL endpoints are considered if measured using validated instruments suited for application in clinical trials ^{227,323}. With regards to uncertainty, the Institute ranks the results of a study according to "high certainty" (randomized study with low bias risk), "moderate" (randomized study with high bias risk), and "low certainty" (non-randomized comparative study). The complete evidence base is then assessed and a conclusion is reached on the probability of the (added) benefit and harm graded according to major added benefit,

considerable added benefit and minor added benefit. Three additional categories are recognized: non-quantifiable added benefit, no added benefit, lesser benefit ^{227,265}.

All types of clinically relevant outcomes are accepted in Sweden, with the inclusion of final outcomes, surrogate endpoints, and composite endpoints, with generic QoL endpoints being preferred over disease-specific endpoints ^{227,273}. Generally, all effects of a person's health and QoL are supposed to be considered as part of the assessment stage, including treatment efficacy and side effects ^{312,313,328}.

In England once again data on all clinically relevant outcomes are accepted with final clinical outcomes (e.g. life years gained) and patient HRQoL being preferred over intermediate outcomes (e.g. events avoided) or surrogate endpoints and physiological measures (e.g. blood glucose levels) ^{243,273,314,320}; particular outcomes of interest include mortality and morbidity. Safety is mainly addressed through the observation of adverse events ³¹⁵. Uncertainty is addressed explicitly through quality of evidence, implicitly, through preference for RCTs and indirectly, through submissions rejection if evidence is not scientifically robust.

Italy, the Netherlands, Poland and Spain include surrogate and composite endpoints in the analysis, in addition to disease-specific quality of life endpoints. Therapeutic value is the most critical criterion for reimbursement in the Netherlands as part of which patient preference data and user friendliness might also be considered ²⁵³.

All countries take into consideration safety data to reflect clinical harm, mainly in the form of adverse events incidence and severity.

Innovation level

In the French setting, clinical novelty is considered by definition through the product's improvement of medical benefit (*ASMR*) relating to the relative added clinical value of the drug which informs the pricing negotiations ²²⁷. Additional innovation characteristics relating to the nature of the treatment (e.g. differentiating between symptomatic, preventive and curative) are also considered but as a second line of criteria ^{227,275,320,328}.

In Germany, clinical novelty is considered implicitly as part of the consideration of added therapeutic benefit for premium pricing. Ease of use and comfort can be reflected indirectly through treatment satisfaction for patients which can be considered as an additional aspect, however not as an explicit factor, similarly to the nature of treatment/technology ³²⁴.

In Sweden, innovation characteristics relating to the added therapeutic benefit (only if it can be captured in the CE analysis) as well as ease of use and comfort are included in the assessment process ^{65,227,320,328}.

As reflected through NICE's operational principles, the encouragement of innovation is an important consideration in England. By definition, the incremental therapeutic benefit as well as the innovative nature of the technology is formally taken into account as part of the product's incremental cost effectiveness ratio (ICER) ³¹⁵.

Among the remaining countries, clinical novelty is essentially considered in all countries; ease of use and comfort might only be considered implicitly and informally if at all, whereas there are mixed approaches in terms of a treatment's technology nature.

Socioeconomic impact

In terms of socioeconomic parameters, in France, 'expected' public health benefit acts as another explicit dimension via an indicator known as public health interest ("Intérêt de Santé Publique", ISP), which is assessed and scored separately by a distinct committee as part of the SMR evaluation but is not used often ^{65,227,256,275}.

In Germany, public health benefit is reflected through the requirement from manufacturers to submit information on the expected number of patients and patient groups for which an added benefit exists as well as costs for the public health system (statutory health insurance) ^{227,324}. All direct costs have to be considered, including both medical and non-medical (when applicable), whereas indirect costs are not a primary consideration but can be evaluated separately if they are substantial, with productivity losses due to incapacity being included only on the cost side ²⁶⁴. In turn, productivity losses due to mortality are only considered in the outcome on the benefit side (to avoid double counting). Budget impact analysis (BIA) is mandatory and should include any one-off investments or start-up costs required in order to implement a new technology, with methodology and sources clearly outlined ^{264,267}.

Among the other study countries, any public health impact of the drug is usually considered but not necessarily in an explicit manner, whereas social productivity is usually reflected through the incorporation of indirect costs, either explicitly or implicitly ²²⁷. In England for example, criteria on social productivity are also considered but not explicitly incorporated as part of cost-effectiveness ratios.

Efficiency

Up until now, cost was not acknowledged as an explicit or mandatory criterion, but BIA, while not mandatory, is highly recommended in France ²²⁷. Although the expert committee had been reluctant to use cost-effectiveness criteria in the evaluation process (^{282,328}, following the bylaw of 2012 (which took effect in 2013) the role of economic evidence was strengthened ²⁵⁹. The CEESP gives an opinion on the efficiency of the drug based on the relative added clinical value (ASMR) of alternative treatments.

In Germany economic analysis (cost-benefit-analysis) is not standard practice in the evaluation but rather can be initiated if no agreement is reached between sickness funds and the manufacturer on the price premium or if the manufacturer does not agree with the decision of the G-BA regarding premium pricing (added benefit); instead, BIA is mandatory (Advance-HTA, 2016). ‘Cost-effectiveness’ acts as one of the most important formal evaluation criteria in Sweden. Parameters having a socioeconomic impact, such as avoiding doctor visits or surgery, productivity impact, and, in general, savings on direct and indirect costs are also considered ³¹².

As already reflected through NICE’s working principles, the relative balance between costs and benefits (i.e. value-for-money) and the effective use of resources should be taken into account in England (e.g. through the explicit cost-effectiveness criterion) ²⁸⁷. Some studies also suggest that the impact of cost to the NHS in combination with budget constraints (budget impact considerations) are taken into account alongside the other clinical and cost-effectiveness evidence ^{131,268,278,282,286}.

In the assessment process by ZIN, the cost-effectiveness criterion follows that of the therapeutic value and the cost consequences analysis. Cost-effectiveness is only considered for drugs with added therapeutic value, which are either part of a cluster and are reimbursed at most at the cluster reference price or are not reimbursed in the absence of possible clustering ^{234,253}. The Netherlands usually performs its own BIA, although voluntary submission from the manufacturer is also an option ^{253,282}.

All other study countries evaluate the efficiency of new drugs through cost-effectiveness evaluation and BIA, but this is not always mandatory.

Other types of evidence

Additional explicit parameters considered in France include the technology’s place in therapeutic strategy mainly in relation to other available treatments (i.e. first-line treatment vs. second-line treatment etc.), and technology’s conditions of use ^{227,258,259}.

Germany is the only country which does not apply any conditions of use in regards to specific sub-populations, in principle reimbursing drugs across the whole indication spectrum as listed on the marketing authorisation ²²⁷.

As reflected through the ethical prioritization framework that the Swedish TLV is using, the ethical considerations of human dignity, need and solidarity are acting as principles for the evaluations.

Beside the notion of clinical need as reflected through NICE's principles, other equity considerations include the 'need to distribute health resources in the fairest way within society as a whole' and the aim of 'actively targeting inequalities', both of which are explicitly mentioned by NICE as principles of social value judgements ²⁸⁷. Equality, non-discrimination and autonomy are other explicit ethical considerations ⁶⁵.

The Netherlands also takes into consideration explicitly ethical criteria based on egalitarian principles, such as solidarity and affordability of the technology by individual patients ^{65,227,271}.

In terms of the remaining countries, conditions for use might be applied for Italy, Poland and Spain, the therapy's place in therapeutic strategy considerations exist for Italy and Spain, whereas ethical considerations are evident in Italy and Poland (implicitly or indirectly), however the use of any additional explicit parameters may not be transparent.

4.3.2.3 Synthesising the evidence and taking into account all factors: Weights

It is not clear how all the factors discussed so far interact with one another, what their relative importance is and what the trade-offs are that HTA bodies are prepared to make between them when arriving at recommendations ^{124,131}. For example, in France, the weights of the assessment parameters considered and the appraisal process overall do not seem to be clear or transparent ³²⁸, although the evidence that informs this judgment is dated and may be contestable. In Spain, the assessment takes into account mainly safety, efficacy, effectiveness and accessibility and it does not consider efficiency and opportunity cost; still the way this is done and the weights of different criteria are not known ²⁴⁵. All countries consider a number of different data sources for the assessment process, with randomised controlled trials (RCTs) usually being the most preferred source for clinical data.

4.3.3 HTA Methods and Techniques Applied

Assuming the existence of an additional benefit (or lesser harm) compared to existing treatment options, all countries are adopting a type of economic evaluation (mainly CUA or CEA) as an analytical method to derive the value of new technologies in addition to efficacy analysis, besides France and Germany, both of which formally used to apply a comparative assessment of clinical benefit as the preferred methodology but with economic evaluation progressively becoming more important as of 2013 but in the context of the existing method of assessment. A summary of analytical methods and techniques applied as part of HTA and their details is presented in Table 4.3.

4.3.3.1 Analytical methods

In Sweden and England the preferred type of economic evaluation is CUA with cost per QALY gained being the favoured health outcome measure, but CEA is also accepted^{237,243,267,272,283,285,287,305,318}. In Sweden CBA with willingness-to-pay (WTP) as an outcome measure can also be applied.

In France, up until now comparative assessment of clinical benefit incorporating final endpoints as an outcome measure acted as the preferred evaluation procedure. However, economic analysis of selected drugs with expected significant budget impact is continuously being considered more formally, especially if its choice is justified and any methodological challenges (especially associated with the estimation of QALYs) are successfully addressed^{65,258,259,267,283,302}. The choice between CEA and CUA depends on the nature of the expected health effects (if expected significant impact on HRQoL then CUA, otherwise CEA).

In Germany, economic evaluations are performed only within therapeutic areas and not across indications, thus an efficiency frontier approach of CBA using patient relevant outcomes is the preferred combination of analysis method-outcome measure^{224,264,267,283,302}. Since the introduction of the AMNOG, economic evaluations are mainly conducted for cases when price negotiations fail after the early benefit assessment and the arbitral verdict is challenged by the technology supplier or the statutory health insurer²⁶⁴.

Table 4.3: HTA methods and techniques applied

Analysis Method								
	FRANCE (HAS / CEESP ⁶⁶)	GERMANY (IQWiG)	SWEDEN (TLV)	ENGLAND (NICE)	ITALY (AIFA)	NETHERLANDS (ZIN)	POLAND (AOTMiT)	SPAIN (RedETS/ISCIII or ICP)
Methods	Comparative Efficacy/ Effectiveness (also CEA, CUA)	CBA but also CUA and CEA (not standard practice)	CUA (also CEA, CBA)	CUA (also CEA, CMA)	CMA, CEA, CUA, CBA ⁶⁷	CEA, CUA, no CMA	Cost-consequences analysis, CEA or CUA - obligatory, CMA (if applicable)	Comparative Efficacy/ Effectiveness, CMA, CEA, CUA, CBA ⁶⁸
Preferred Outcome measure	Final outcome, Life Years (QALY, if CUA; Life Years, if CEA)	Patient relevant outcome (can be multidimensional) - efficiency frontier (QALY)	QALY (WTP, if CBA)	QALY (cost per life year gained, if CEA)	Final outcome, Life Years (QALY, if CUA or CEA; Life Years, if CEA)	Effectiveness by intention-to-treat principle, and expressed in natural units - preferably LYG or QALY	QALY or LYG	QALY in CUA
Utility scores elicitation technique	EQ-5D and HUI3, from general French population	Utility scores from patients, Direct (e.g. TTO, SG), Indirect (EQ-5D)	Utility scores from patients, Direct (e.g. TTO, SG), Indirect (EQ-5D)	Utility scores from general English population, Direct (e.g. TTO, SG), Indirect (EQ-	Both direct and indirect (EQ-5D) elicitation techniques	Either Direct (TTO, SG, VAS), or Indirect (EQ-5D); selection should be justified	Direct or indirect utility scores ⁶⁹	Utility scores from general Spanish population, Direct (e.g. TTO, SG), Indirect (EQ-5D) ⁷⁰

⁶⁶ In France, economic evaluations are undertaken only of select drugs with expected significant budget impact.

⁶⁷ A template for the submission of the P&R dossier to AIFA is in progress.

⁶⁸ For the case of drugs at central level carried out by ICP, comparative efficacy/effectiveness is taken into account. The ICP receives the so called “Informe de Posicionamiento Terapéutico” (Therapeutic Positioning report), a therapeutic assessment conducted by the Spanish Medicines Agency (Agencia Española del Medicamento) based on which confidential discussions around the appraisal of the drugs takes place but which does not take into consideration cost-effectiveness. Economic evaluations are mainly taking place for the case of non-drug technologies under the scope of RedETS.

⁶⁹ It is recommended to use indirect methods for preferences measurement – validated questionnaires in Polish. While measuring preferences with the EQ-5D questionnaire, it is advised to use the Polish utility standard set obtained by means of TTO.

⁷⁰ Surveys or previously validated HRQOL patient surveys.

				5D), Systematic review				
Comparator	Usually 'best standard of care' but can be more than one ⁷¹	Usually 'best standard of care' but can be more than one ⁷²	Usually 'best standard of care' but can be more than one ⁷³	Usually 'best standard of care' but can be more than one ⁷⁴	Usually 'best standard of care' but can be more than one ⁷⁵	Treatment in clinical guidelines of GPs; if not available, most prevalent treatment	'Best standard of care' which is reimbursed in Poland ⁷⁶	Best standard of care, usual care and/or more cost-effective alternative
Perspective	Widest possible to include all health system stakeholders ⁷⁷	Usually statutory health insurant ⁷⁸	Societal	Cost payer (NHS) or societal if justified	Italian National Health Service ⁷⁹	Societal (report indirect costs separately)	The public payer's perspective, public payer + patient (by law)	Cost payer (NHS) and societal (rarely used), and they should be presented separately
Subgroup analysis	Yes (when justified)	Yes	Yes	Yes	Yes	Yes	Yes (if needed, but decreases validity)	Yes
Clinical Evidence								
Preferred study design	Head-to-head RCTs; other designs accepted if no RCTs available	Head-to-head RCTs; other designs accepted in the absence of RCTs	Head-to-head RCTs; other designs accepted if no RCTs available	Head-to-head RCTs; other designs accepted if no RCTs available	Head-to-head RCTs; other designs accepted if no RCTs available	Head-to-head RCTs	Head-to-head RCTs; other designs accepted if no RCTs available.	Head-to-head RCTs; other designs accepted if no RCTs available

⁷¹ Including most cost-effective, least expensive, most routinely used, and newest.

⁷² Including most cost-effective, least expensive, and most routinely used. If the efficiency frontier approach is used as part of CBA, then “all relevant comparators within the given indication field” must be considered.

⁷³ Including most cost-effective, least expensive, and most routinely used.

⁷⁴ Including most cost-effective, least expensive, and most routinely used.

⁷⁵ Including most cost-effective, and most routinely used.

⁷⁶ These might include a) most frequently used; (b) cheapest; (c) most effective; and (d) compliant to the practical guidelines.

⁷⁷ Needs justification (especially if societal).

⁷⁸ Also community of statutorily insured, perspective of individual insurers, or the societal perspectives are possible.

⁷⁹ Societal perspective is not mandatory but it can be provided in separate analysis.

Systematic literature reviews for collecting evidence required/conducted by regulator	Yes, guidelines provided/Yes, in French	Yes / No	Not mandatory	Yes / Yes	Yes / Yes	Yes / Yes	Yes	Not always ⁸⁰
Meta-analysis for pooling evidence	Not specified	Not specified for new drugs	Not specified	Yes	Yes	Yes, encouraged	Yes	No ⁸¹
Data extrapolation	Qualitative only, in absence of effectiveness data form RCTs	No	Quantitative, both in absence of RCT effectiveness data and in absence of long-term effects	Qualitative and quantitative, both in absence of RCT effectiveness data and in absence of long-term effects	Quantitative. Qualitative in absence of RCT effectiveness data	Qualitative, in the absence of RCTs and in absence of long-term effects	Possible if needed but not recommended	Quantitative, in the absence of effectiveness data
Resources/costs								
Types	Direct medical, direct non-medical, indirect (both for patient and carer)	Depending on perspective: direct medical, informal costs, productivity loss (as costs)	Direct medical, direct non-medical, indirect (both for patient and carer)	Direct medical, social services	Direct costs only. Indirect costs can be taken into account in a separate analysis	Both direct and indirect costs inside and outside the healthcare system	Direct medical costs, direct non-medical costs	Direct and indirect costs (on rare occasions), costs of labour production losses or lost time, informal care costs

⁸⁰ For non-drugs under RedETS, systematic literature review is always conducted.

⁸¹ For non-drugs under RedETS, meta-analysis might be conducted.

Data source/Unit Costs	Direct : PMSI (Programme de Médicalisation des Systèmes d'Information) Indirect: human capital costing, friction costing	Statutory health insurance, further considerations depending on perspective chosen	Drugs: pharmacy prices Indirect: human capital costing	Official DoH listing	Variety of sources ⁸²	Reference prices list should be used	Variety of sources ⁸³	Official publications, accounts of health care centres, and the fees applied to NHS service provision contracts
Discounting								
Costs	4.0% (up to 30 years) and 2.0% after	3.0%	3.0%	3.5%	Not available (update in progress)	4.0%	5.0%	3.0%
Outcomes	4.0% (up to 30 years) and 2.0% after	3.0%	3.0%	3.5%	Not available (update in progress)	Under review - will probably be set at same level as costs discounting	3.5%	3.0%
Sensitivity Analysis	0% , 3.0% (6% max)	0% to 5%	0% to 5.0%	0% to 6.0%	Not available (update in progress)	Not obligatory	5% and 0% for costs and outcomes 0% for outcomes 5% for costs ⁸⁴	From 0% to 5.0%
Time Horizon								
Time Horizon	Long enough so that all treatment	At least the average (clinical) study duration; longer for chronic conditions,	Time needed to cover all main outcomes and costs	Long enough to reflect any differences on outcomes and	Duration of the trial is considered ⁸⁵	Primarily based on duration of RCTs ⁸⁶	Long enough to allow proper assessment of differences in health outcomes and costs	Should capture all relevant differences in costs and in the

⁸² Prices available in the Official Journal of the Italian Republic (Gazzetta Ufficiale), accounts of health care centres, the fees applied to NHS service, scientific literature/ad hoc studies.

⁸³ Including a) list of standard costs, b) formerly published research, c) local scales of charges, d) direct calculation.

⁸⁴ It is currently under revision (AOTMiT HTA Guidelines updating process) so may change soon.

⁸⁵ Additional long term evidence collected through monitoring registries.

⁸⁶ Secondary horizons include any longer needed depending on the context of interest.

	outcomes can be included	especially if lifetime gains are expected; same horizon for costs and benefits		costs between technologies compared			between the assessed health technology and the comparators	effects of health treatments and resources ⁸⁷
Thresholds								
Thresholds	No threshold (only eligibility threshold to conduct economic evaluation)	Efficiency frontier (Institute's own approach)	No official threshold; 50% likelihood of approval for ICER between €79,400 and €111,700	Implicit: ~ £20-000 - £30,000 per QALY; Empirical: £12,936 per QALY	No threshold in use	No official threshold	3 x GDP per capita for ICUR(QALY) or ICER(LYG)	Unofficial: €21,000 – €24,000 /QALY (Recently provided by SESCS ⁸⁸ to the Spanish MoH)

⁸⁷ In some cases, the time horizon will have to be extended to the individual's entire life.

⁸⁸ Servicio de Evaluación y Planificación, Canarias.

In the Netherlands and Italy the preferred type of economic evaluation is CUA if the improvement in quality of life forms an important effect of the drug being assessed, or, if this is not the case, a CEA ^{239,246}. In Spain, any of the four methods of analysis may be used (CMA, CEA, CUA or CBA).

4.3.3.2 Types of clinical evidence considered

In relation to clinical evidence, all countries acknowledge that randomized controlled head-to-head clinical trials is the most reliable and preferred source of treatment effects (i.e. outcomes), with data from less-rigorous study designs being accepted in most study countries (England, France, Germany, Sweden, Poland, Spain, Italy) e.g. when direct RCTs for the comparators of interest are not available ^{283,315,320}.

Most agencies require systematic literature reviews to be submitted by manufacturers as a source of data collection and carry out their own reviews. A meta-analysis of key-clinical outcomes is recommended for pooling the results together given the homogeneity of the evidence in England, Italy, Netherlands and Poland ^{283,315}.

If evidence on effectiveness is not available through clinical trial data, then France and Netherlands allow for a qualitative extrapolation based on efficacy data, with Spain conducting quantitative extrapolation and Sweden, England, Italy and Poland applying both qualitative and quantitative modelling. In Sweden, England and Netherlands, short-term clinical data are extrapolated also if data on long-term effects are absent.

4.3.3.3 Resources/cost evidence

In terms of resources used, in addition to direct medical costs, France and Sweden consider all relevant costs including direct non-medical and indirect costs, both for patients and carers ^{267,283}; however, only direct costs are considered in the reference case analysis and incorporated in the ICER in the case of France ²⁵⁸. Germany also takes into account informal costs and productivity gains separately as a type of benefit, whereas England additionally considers cost of social services.

Poland incorporates direct medical costs and direct non-medical costs. In the Netherlands, the Health Care Insurance Board's "Manual for cost research" applies for the identification, measurement and valuation of costs; pharmacoeconomic evaluations need to include both direct and indirect costs inside and outside the healthcare system ²⁴⁶. In Italy it

is recommended to include direct costs; indirect costs can be taken into account in a separate analysis ²²⁷. Spain incorporates both direct and indirect costs (the latter on rare occasions) as well as costs of labour production losses or lost time and informal care costs in the analysis ^{227,302}. Finally all countries recommend the application of country-specific unit costs ²⁸³.

4.3.3.4 Discounting and time horizon

In all study countries both costs and benefits are discounted ^{267,302,318,320}, and uncertainty arising due to variability in model assumptions is investigated usually in the form of sensitivity analysis. Generally, in each country discounting levels for costs are equal to discounting levels for outcomes, ranging from 3% to 5%, with the complete range of 0% to 5% (or even 6%) usually being explored with sensitivity analysis. In Italy, information on discounting is not available at the moment given an update in progress by AIFA ²²⁷. In terms of a suitable time horizon, none of the countries uses an explicit time frame but instead they adopt a period that is long enough to reflect all the associated outcomes and costs of the treatments being evaluated, including the natural course of the disease ^{262,267}.

4.3.3.5 Acceptable “value for money” thresholds

No explicit, transparent, or clearly defined cost-effectiveness thresholds exist in any of the countries except for Poland, Spain and, possibly, England.

In line with the World Health Organization (WHO) suggestions of two to three times GDP per capita, a three times GDP per capita threshold has been implemented in Poland. Generally, a drug is deemed cost-effective by AOTMiT if estimates are greater than one time the GDP per capita, but smaller than 70,000 PLN per QALY/LYG ²⁹². In Spain, a €21,000 - €24,000 per QALY threshold was recently provided by Servicio de Evaluación y Planificación Canarias (SESCS) to the Ministry of Health ²²⁷.

In England, although some evidence suggests the existence of an implicit threshold ranging somewhere between £20,000 and £30,000 ^{231,272,295,314}, it is evident that such a range is not definite mainly because some products with a cost per QALY below these ranges receiving negative coverage decisions and other products above these ranges ending up with positive recommendations ^{243,288,304}. Indeed, several studies point towards the existence of a threshold range based on which additional evidence on several factors are required for the recommendation of technologies with an ICER of above £20,000, and even stronger evidence of benefit in combination with explicit reasoning required for the coverage of

technologies with an ICER above £30,000 ^{231,237,268,306,315,328}. However, a more recent study using data on primary care trust spending and disease-specific mortality estimated an empirical based “central” threshold of £12,936 per QALY, with a probability of 0.89 of less than £20,000 and a probability of 0.97 to be less than £30,000 ³²².

In Germany, the efficiency frontier approach is used to determine an acceptable “value for money”, even though this is not involved in the process of the initial rebate negotiations. In Sweden, recent evidence suggested that the likelihood of approval is estimated to be 50% for an ICER between €79,400 and €111,700 for non-severe and severe diseases respectively ³²⁹. In the Netherlands, there is no formal threshold in place but there have been some attempts to define one. The €20,000 per life-year gained (LYG) threshold used in the 1990s to label patients with high cholesterol levels eligible for treatment with statins has been mentioned in discussions on rationing, but was never used as a formal threshold for cost-effectiveness. The same was the case with a threshold that the Council for Care & Public Health wanted to implement based on criteria such as the gross domestic product (GDP) per capita, in line WHO recommendations, which for the Netherlands would translate into €80,000/ QALY ²³⁴. The Council also suggested that the cost per QALY may be higher for very severe conditions (a tentative maximum of €80,000) than for mild conditions (where a threshold of €20,000 or less may be applied) ²³², but none of the above was ever implemented.

4.3.4 HTA Outcomes and Implementation

In all countries, assessment and appraisal of outcomes are mainly used as a tool to inform coverage decisions relating to the reimbursement status of the relevant technologies; all countries use the results to inform pricing decisions directly or indirectly. A summary of the types of HTA outcomes and implementations in the study counties is presented in Table 4.4.

Table 4.4: HTA outcomes and implementation

		FRANCE (HAS / CEESP)	GERMANY (IQWiG)	SWEDEN (TLV)	ENGLAND (NICE)	ITALY (AIFA)	NETHERLANDS (ZIN)	POLAND (AOTMiT)	SPAIN (RedETS/ISCIII or ICP)
Publicly Available Report		Yes, both in French and English ⁸⁹	Yes	Yes (summary report with some details on cost-effectiveness)	Yes	Yes, in the Official Journal of the Italian Republic (Gazzetta Ufficiale)	Yes	Yes (in Polish on the AOTMiT website), but confidential information is publicly unavailable	No for drugs ⁹⁰
Policy Implication	Reimbursement	Yes, through SMR ⁹¹	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Pricing	Yes, through ASMR ⁹²	Yes	Yes	No, only indirectly as it has an impact on product's ICER	Yes	Yes, except certain expensive medicines ⁹³	Yes, if reimbursement decision is positive	Yes
Access restrictions		Yes, various restrictions in place ⁹⁴	Existence of managed entry agreements but details not	Yes, restrictions for specific subpopulations, temporary decisions and	Yes, major and minor restrictions as well as performance	Yes, various managed entry agreements ⁹⁵	Yes, system of coverage with evidence development (CED)	Yes, including major and minor ⁹⁶	Yes

⁸⁹ Economic Evaluation reports are available but some parameters are deleted in the public version (elements related to medicines costs mainly).

⁹⁰ For non-drug technologies under RedETS usually yes in the form of bulletins and web pages of HTA agencies.

⁹¹ The level of SMR determines if a drug shall be reimbursed and if yes, at which level (low 15%, mod 30%, high 65%).

⁹² The level of ASMR is used for pricing negotiations with manufacturers.

⁹³ A bureau of the government on a case-by-case approach negotiates rebates with the industry for certain expensive medicines (actual price is 'secret' but hospitals can ask for an add-on).

⁹⁴ Including recommendation to only reimburse this medicine in second intention, restrictions to specific sub-populations, Financial risk-sharing (price-volume agreements and budget caps).

⁹⁵ Such as PVAs, cost-sharing, budget cap, monitoring registries, payment by results, risk-sharing, therapeutic plans, and "AIFA notes".

⁹⁶ Major include restricted to specific subpopulations (monitoring of use); Minor include requiring a lower price so called Risk Sharing Schemes (cost sharing in practice).

		publicly available	risk sharing agreements	based agreements				
Dissemination	Publicly available online	Dossier assessment, Reports, Rapid reports, Addendums	Informational material distributed to the major stakeholders, decisions published online	Publicly available online	Monthly AIFA publication of price lists of reimbursed products. Annual publication of data on pharmaceutical expenditure and consumption (Rapporto Osmed)	Online for general public and distributed to stakeholders	Publication online	No for drugs ⁹⁷
Implementation	Prescription guidelines, drug formularies and positive list	Prescription advice issued by G-BA based on therapeutic assessment ("Therapiehinweise")	Drug formularies	Prescription guidelines, drug formularies	A product can be assigned to Class A, H or C ⁹⁸	Positive list. In case of therapeutic equivalent, the drug is either not accepted for public reimbursement or subject to a reference pricing system	Different reimbursement lists categories ⁹⁹	Inclusion in the national reimbursement list
Appeal	Yes ¹⁰⁰	Yes, through arbitration board ¹⁰¹	Yes	Yes	Companies can appeal to Court but there is no specific appeal procedure	Yes	No	Yes

⁹⁷ RedETS reports for non-drugs become publicly available.

⁹⁸ Class A refers to products reimbursed by the NHS. Class H refers to products for hospital use. Class C refers to non-reimbursed products.

⁹⁹ Pharmacy drugs (Rx drugs; 30% or 50% patient co-payment, lump sum, no co-payment); drug programmes (selected diseases and patients; free); chemotherapy drugs (hospital settings; free); drugs reimbursed in off-label indications.

¹⁰⁰ Manufacturers can appeal to decisions made by both commissions. They are then called for an audition to explain their position.

¹⁰¹ Manufacturers have the right to commission CBA if they do not agree with the established added benefit.

Revision	Yes, every 5 years or sooner if decision from HAS or request from the MoH.	Yes, at least one year after benefit assessment ¹⁰²	Yes	Yes	Yes ¹⁰³	Yes, but not on a regular basis ¹⁰⁴	Yes, 2 years after first assessment, 3 year after 2 nd , 5 years after 3 rd assessment	Yes
----------	--	--	-----	-----	--------------------	--	--	-----

¹⁰² In some cases, decisions are time-limited, revision takes place once the term is over.

¹⁰³ The negotiation process leads to a 2 year confidential, renewable contract between AIFA and the manufacturer.

¹⁰⁴ In practice, providers that have no adequate reimbursement due to a new innovation will ask the Dutch healthcare authority for a revision of reimbursement. The agency then investigates if a revision is reasonable and what the new reimbursement should be.

4.3.4.1 Timing and public availability

Generally the time needed for the evaluation of a health technology to be completed differs from country to country. However, in line with the EU Transparency Directive, all countries must have reached a decision on pricing and reimbursement within 180 days post marketing authorisation ³²⁸. In all countries the final decision report is publicly available, usually through the HTA agency's website ^{216,328}, and the policy implication of the evaluation outcome relates to the pricing and reimbursement status of the technology: reimbursement (List), no reimbursement (Do Not List), or conditional reimbursement (List with Restrictions) ^{278,328}.

4.3.4.2 Policy implications

In France and Sweden, only drugs with additional therapeutic value can “obtain a higher reimbursement basis” ³²⁸; in France, by assessing the evidence of the product's medical benefit or medical service rendered (Service Médical Rendu, SMR), the improvement in medical benefit and added therapeutic benefit (Amélioration du Service Médical Rendu, ASMR) are derived, which determine the reimbursement status and influence the price level of the product respectively, whereas in Sweden the outcome of the evaluation can also drive the price setting in addition to coverage decisions ^{312,313}. In Germany, the outcome of the clinical/economic evaluation will be used mainly to inform the negotiation between sickness funds and manufacturer on the price premium. In England reimbursement status has no direct effects on price, but indeed price indirectly affects the reimbursement status of the drug as it will have an impact on the ICER. In the Netherlands, the positive outcome of an HTA results to the inclusion of the medical technology in the positive list ²⁵³. In terms of the reimbursement decision, if the cost-effectiveness analysis for a new innovative drug is of good quality, reimbursement will principally not be denied on the basis of cost-effectiveness, despite potentially relatively high cost-per-QALY values ²³⁴. Finally, in Italy, if a reimbursement status is approved, the pricing is decided simultaneously. If the reimbursement decision is negative, the product will be put on the negative list and price is determined by the manufacturer (“free pricing”).

4.3.4.3 Access restrictions

All countries apply access restrictions usually relating to specific indications or specific population sub-groups. France mainly uses financial risk-sharing (price-volume) agreements³²⁸. Sweden issues temporary decisions for cases when there is insufficient certainty around the (clinical) evidence³²⁸ and risk sharing agreements may take place to speed up the reimbursement process upon the requirement of additional evidence ex post the review²⁹⁴, in addition to restricting access for specific sub-populations. In England, major and minor restrictions exist: the former relate to cases where the technology is indicated only for second-line treatment (and beyond) or for only specific sub-population, and the latter relate to the need for specialist supervision or treatment monitoring²⁶⁸; performance based agreements (or response rules) also exist, especially in regards to the use of biologics and cancer drugs, according to which a pre-specified clinical (endpoint) condition must be reached at a specific post-assessment time point for the coverage of the technology to continue²³⁵. In the Netherlands, the system of coverage with evidence development (CED) for high cost and orphan inpatient drugs has been extensively used between 2006 and 2011. Currently, financial-based agreements and performance-based risk sharing agreements are considered as well. In Poland, restrictions could be applied to a positive recommendation, which can be either major, e.g. restricted to specific subpopulations (monitoring of use), or minor: e.g. requiring a lower price (so called Risk Sharing Schemes, but cost sharing in practice)²²⁷. In Spain, MEAs are concluded at the regional level. PVAs agreements are usually applied to single new products where the negotiated price is conditioned by the expected number of units sold.

4.3.4.4 Dissemination and implementation

Most countries employ dissemination procedures in order to support the implementation of their decisions, including prescribing guidelines and national drug formularies²⁵³. In France, since 2013, there is a public online drug database allowing the general public to access data and documents on marketed drugs²⁶⁰. In Germany, IQWiG prepares a variety of dissemination products besides the dossier assessment including technical scientific reports (and rapid reports where no commenting procedures take place) but also public and user-friendly health information and working papers on recent developments on the field and methodological aspects²⁶⁵. The dossier assessment is provided by the G-BA which can also issue prescription advice²²⁷. In Sweden, at least for the review of products that are already

on the positive list, informational material in the form of a fact sheet is produced (possibly accompanied by supplementary information taking the form of a PowerPoint presentation and an FAQ sheet), covering the analysis, the appraisal and the conclusion of the evaluation, distributed to the major stakeholders on the date of the decision and about a week before it becomes publicly available online ^{312,313}. In England, NHS is legally obliged to implement NICE guidance and fund the recommended technologies within 3 months from the outcome of the decision, possibly by displacing resources from the use of other technologies ^{243,315}. In Poland, since the Reimbursement Act (issued in 2011, effective from January 1st, 2012), drugs can be reimbursed under different lists ²²⁷. Pharmacy reimbursement includes prescribed-only medicines available to patients through four main categories of co-payment. Chemotherapy drugs are available in hospital setting free of charge. Other “regimen” programs are available under which drugs for selected diseases are reimbursed free to strictly defined patient populations whose eligibility is decided from a clinicians’ committee.

4.3.4.5 Appeal mechanisms and review of decisions

Most countries have appeal mechanisms in place in case of dissent, and they all revise their decisions either according to fixed time schedule or on a rolling basis ^{320,328}; in France, the drug registration is subject to renewal every 5 years and a drug may also be subject to post-registration studies. Sweden re-evaluates its old reimbursement list, and both Sweden and England may revise technologies once new evidence becomes available. On average the positive recommendations (with or without restrictions) are around 90% for NICE ²⁴¹. Although it appears that revisions were taking place systematically after four years for in-patient drugs and on an *ad hoc* basis for out-patient drugs ^{242,328}, more recent evidence suggests that in practice, the process is irregular and providers that have no adequate reimbursement due to a new innovation will ask the Dutch healthcare authority for a revision of reimbursement. The agency then investigates if a revision is reasonable and what the new reimbursement should be ²²⁷. In Italy, the negotiation process leads to a two year confidential, renewable contract between AIFA and the manufacturer ²²⁷, and a possible revision is feasible on the grounds of a new product exceeding the original forecast of a company.

4.4 Discussion

In all study countries HTA agencies are autonomous bodies. The evaluation process of medical technologies includes an initial assessment of evidence conducted by technical groups, followed with the appraisal of the assessed evidence from an expert committee that is producing reimbursement and coverage recommendation(s) for the final decision body which can be either the payer (e.g. MoH, HIF), or the HTA agency itself.

In addition to the comparative assessment of clinical benefit, most countries implement a type of economic evaluation (mainly CUA or CEA) as the main analytical method to determine the value of new technologies, with the preferred health gain measure usually being the QALY or other alternative patient-relevant (if not final) outcomes. For the elicitation and measurement of utility scores, preference-based elicitation techniques (e.g. time trade-off, standard gamble) are used to elicit utility scores either from patients or the general population.

The evaluation (assessment and appraisal) outcome is mainly used as an aid to make coverage decisions in relation to the reimbursement status of the medical technologies, but in all countries the analysis outcomes are used to influence pricing decisions as well (indirectly in England). Access restrictions for sub-populations or sub-indications, possibly through the application of risk sharing agreements, have become common practice across many jurisdictions. Information material is often disseminated by the HTA agencies to a range of stakeholder groups; the implementation of the agencies decisions is usually taking the form of prescription guidelines and drug formularies. Technology suppliers across all jurisdictions have the option of dissent/appeal and revision of the decision is taking place either every a standard period of time or when new evidence becomes available.

The results show that different additional criteria beyond economic evaluation or clinical benefit assessment are captured or included in the evaluation process that may explain heterogeneity in coverage recommendations and decision-making.

Overall, all countries assess similar types of evidence, however the specific endpoints used, their level of provision and requirement, the way they are incorporated (e.g. explicitly vs. implicitly) and their relative importance varies across countries. The main evidence assessed could be divided into six clusters of information: (a) burden of disease, (b) therapeutic & safety impact, (c) innovation level, (d) socioeconomic impact, (e) efficiency considerations and (f) other sources of evidence and criteria.

4.4.1 Conceptual and Methodological Limitations in Value Assessment

Current value assessment approaches mainly consider (comparative) clinical efficacy in combination with clinical cost-effectiveness techniques, while increasingly incorporating real world evidence following drug market entry, thus reflecting comparative effectiveness and efficiency. However, there is considerable subjectivity in the criteria selection used to interpret evidence and determine product value, often lacking a sound basis and therefore leading to arbitrariness, for example relating to which metrics to use for measuring efficacy and effectiveness, what type of costs to consider, and very importantly how to account for other key dimensions of value.

Most of the value assessment approaches examine the efficacy/effectiveness, or cost-effectiveness of new interventions by mostly addressing only a partial dimension of ‘overall value’ in a systematic and explicit manner that mainly relates to ‘Scientific Value Judgments’ (ScVJ) of their therapeutic aspect (e.g. safety, efficacy, effectiveness), possibly in relation to cost. However, the value of new medical technologies is multi-dimensional and not only limited to clinical benefit and cost. In addition to commonly used ScVJ which is based solely on “scientific” evidence relating to clinical cost-effectiveness and ICERs, other “social” value factors falling under the information clusters of burden of disease, innovation level and socioeconomic impact also play a definitive role in decision-making through the exertion of ‘Social Value Judgements’ (SoVJ), however they are rarely formally incorporated in the evaluation process.

In most settings, the absence of clarity on the use of SoVJ, including their interplay with ScVJ and their influence on decisions remains unknown. Social value elements are usually considered implicitly by HTAs or decision-makers, mostly non-transparently on an *ad hoc* basis. For most of them it is not known what their relative importance is, and what the trade-offs are that HTA bodies are willing to make. As a result, the definition of value could be regarded as an elusive concept given that a multiplicity of evaluation criteria applies across different settings and with differential intensity in a non-systematic manner.

4.4.2 Policy Implications and Ways Forward

Following the technical review of policy initiatives and opportunities for collaboration and research for access to new medicines in Europe, WHO proposes for more extensive use of HTA in decision-making³³⁰. However for this to take place a more holistic perspective and coordinated actions would be needed.

Decision-makers as well as other stakeholders need clear, comprehensive and transparent ways of assessing clinical and economic benefit and the impact those new treatments have from a wider socio-economic perspective in order to make rational decisions about priority setting. Not having such methods creates a conceptual, methodological and policy gap. Appropriate adaptations on the current methodologies or development of new transparent conceptual frameworks seem to be needed.

NICE in England is one of the forerunner agencies in acknowledging, formalising and creating a methodological landscape for SoVJ, which include, first, the burden of disease the treatment addresses, hence the clinical and policy importance of the health topic under consideration; second, the cost impact on resources from a societal perspective; third, policy objectives relating to the long-term benefits of innovation¹⁴⁶⁻¹⁴⁸, and, in general, the broader balance between benefits and costs. The existing influence of disease severity could be illustrated from the case of end of life treatments, where QALYs gained for terminal illnesses are having a greater weight¹⁵³, on the grounds that society places a special value on extending the lives of the terminally ill⁶⁷. Decision-makers have been exploring new ways of considering additional value parameters, while highlighting the need for “a broader and more transparent assessment” methodology, suggesting a move towards value based assessment^{31,154}.

Aspects of HTA shortcomings have also been reflected by various recent initiatives seeking to establish “value frameworks” aiming to aid pricing and clinical practice decisions by considering a variety of parameters for the assessment of value, possibly in relation to costs. A lot of that work has been driven by health care professional associations^{170-172,174,175}. However, attention should be paid on their methodologies, for recommendations to be robust and avoid misguided decisions¹⁷⁷. All these initiatives have attempted to adopt multi-criteria evaluation approaches, albeit in a very simplified and relatively abstract manner. Other approaches embedded in decision analysis could address benefit-risk assessment considerations of health care interventions^{72,331}. Considering the limitations this systematic review has highlighted in the context of HTA as it is practised currently, it looks as though multi-criteria decision analysis methods could be explored to capture the value of new medical technologies in a holistic manner and, through this, facilitate HTA decision-making processes in a spirit of transparency, encompassing and robustness^{332,333}.

4.5 Conclusion

The study highlights a number of significant similarities but also considerable differences in practices, processes and policies of value-assessment for new medicines across the study countries. These differences could exist because of different national priorities between countries, but also because of different processes and methodological frameworks adopted for the elicitation of decision-makers' preferences. Overall, there is significant uncertainty with regards to what additional value criteria to incorporate, how to establish their relative importance, and whose preferences to consider. Currently, all these decisions are subject to DMs' discretion but are, in most cases, exemplified in a less than transparent way, potentially resulting in some form of bias.

Procedures characterized by greater transparency or clarity in terms of value criteria used and a higher degree of encompassing and methodological robustness could lead to more rational evidence-based decision-making, contributing to more efficient resource allocation and, potentially, higher societal welfare, while also raising public confidence and fairness in terms of homogeneity and consistency of decision outcomes.

The limitations of the current value assessment methodologies and the identified conceptual and policy gaps suggest that there is a need for alternative methodological approaches that encompass multiple evaluation criteria explicitly, so that value can be an explicit function of a number of parameters beyond those currently used. This is increasingly becoming imperative in the context of European collaboration, particularly if some form of joint assessment at EU level is likely to emerge beyond 2020. Decision analysis and multi-criteria evaluation approaches could provide the foundation for alternative ways of measuring and eliciting value of new medicines and technologies as they provide a comprehensive alternative for quantitative modelling.

Chapter 5 – Paper 3

Methodological Framework: a Multiple Criteria Decision Analysis Value Measurement Model and Techniques for the Evaluation of New Medicines in Health Technology Assessment¹⁰⁵

Summary

Escalating drug prices have catalysed the generation of numerous “value frameworks” with the aim to inform payers, clinicians and patients around the assessment process of new medicines for the purpose of coverage and treatment selection decisions. Although this is an important step towards a more inclusive Value Based Assessment (VBA) approach, aspects of these frameworks are based on weak methodologies and could potentially result in misleading recommendations or decisions.

A Multiple Criteria Decision Analysis (MCDA) methodological process based on Multi Attribute Value Theory (MAVT) is adopted for building a multi-criteria evaluation model. A five-stage model-building process is followed, using a top-down “value-focused thinking” approach, involving literature reviews and expert consultations. A generic value tree is structured that captures decision-makers’ concerns for assessing the value of new medicines in the context of Health Technology Assessment (HTA) and in alignment with decision theory.

The resulting value tree (Advance Value Tree) spans three levels of criteria (top level criteria clusters, mid-level criteria, bottom level sub-criteria or attributes) relating to five key domains that can be explicitly measured and assessed: (a) burden of disease, (b) therapeutic impact, (c) safety profile (d) innovation level, and (e) socioeconomic impact. A combination of MAVT modelling techniques is proposed for operationalising (i.e. estimating) the model: an indirect elicitation technique for value functions based on pairwise qualitative judgments for scoring the alternative options (MACBETH), an indirect qualitative swing weighting technique for assigning relative weights of importance to the criteria (MACBETH-

¹⁰⁵ Chapter 5 has been published with co-author Prof Panos Kanavos as: Angelis A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for Evaluating New Medicines in Health Technology Assessment and Beyond: the Advance Value Framework. *Social Science & Medicine* (2017). In press; doi.org/10.1016/j.socscimed.2017.06.024

weighting), and a simple additive aggregation technique for combining scores and weights together.

Overall, the combination of these MCDA modelling techniques for the elicitation and construction of value preferences across the generic value tree provides a new value framework (Advance Value Framework) enabling the comprehensive measurement of value in a transparent and structured way. Given the fully flexibility to meet diverse requirements and become readily adaptable across different settings, it enables its use as a decision-support tool for policy-makers around the coverage and reimbursement of new medicines.

5.1 Background

Scarce resources, rising demand for health services, ageing populations and technological advances threaten the financial sustainability of many health care systems and render efficient and fair resource allocation a cumbersome task ^{4,6-9}. Decision-making in health care is inherently complex as numerous objectives need to be balanced, usually through the involvement of many stakeholders. One set of tools used widely to improve efficiency in resource allocation is Health Technology Assessment (HTA). The use of HTA has expanded significantly over the past 20 years and is used to assess and appraise the value of new medical technologies as well as inform coverage decisions.

Evidence-based medicine ^{334,335}, economic evaluations ^{336,337}, burden of disease estimates ³³⁸, and budget impact analysis ²⁸² can be used to inform decisions on resource allocation. Nevertheless, they offer limited guidance to decision makers, as their results cannot be integrated and judged simultaneously and neither can the value trade-offs associated with them ⁶⁴. The use of economic evaluation techniques such as cost effectiveness analysis (CEA) has become the preferred analytical method adopted by many HTA agencies. However, not all the value concerns of decision-makers for evaluating the performance of different health care interventions are adequately reflected in a cost effectiveness model ³³⁹. For example, the use of cost utility analysis (CUA) and the cost per unit of quality adjusted life year (QALY) has become the metric of choice by many HTA agencies when assessing and appraising value. By definition though, it only considers length of life in tandem with health related quality of life, and does not adequately capture social value such as the wider innovation and socioeconomic impact ^{209,210}.

Due to the complexity of these multiple criteria problems, decision-makers tend to adopt intuitive or heuristic approaches for simplification purposes, but as a consequence important information may be under-utilised or be altogether excluded leading to choices

based on an *ad hoc* priority setting process⁶⁴. Eventually the decision making process tends to be explicitly informed solely by evidence from economic evaluations, with social value concerns being considered on an implicit and *ad hoc* basis. As a consequence, when faced with multiple trade-offs across a range of societal values, decision makers seem not to be well equipped to make informed and rational decisions^{70,71}, therefore diminishing the reasonableness and credibility of the decision outcomes. It is probably the case that a more “rational” approach is needed that can simultaneously take into account the multiplicity of criteria, and that can aggregate the performance of alternative interventions across the criteria of interest while accounting for differences between their relative importance, therefore enabling the overall construction and analysis of decision makers’ preferences in a simple and transparent way. The multiple initiatives that have emerged over the past few years through the development of value frameworks aiming to aid reimbursement agencies, health care professionals and patients understand the value of new therapies and make better choices about their use serve as a testament to this particular gap.

Some of the most prominent and well known value frameworks that have attracted a lot of attention include those proposed by the American College of Cardiology and the American Heart Association (ACC/AHA)¹⁷⁰, the American Society of Clinical Oncology (ASCO)¹⁷¹, the European Society of Medical Oncology (ESMO)¹⁷², the Institute for Clinical and Economic Review (ICER)¹⁷³, the Memorial Sloan Kettering Cancer Centre (MSKCC)¹⁷⁴, the National Comprehensive Cancer Network (NCCN)¹⁷⁵, and the Working Group on Mechanisms of Coordinated Access to Orphan Medicinal Products (MoCA-OMP)¹⁷⁶ among others. These value frameworks adopt multiple criteria approaches in an attempt to decompose complex problems into slightly simpler ones and address these sequentially. As such, they are an important step towards a more inclusive Value Based Assessment (VBA) process despite being perceived as weak and atheoretical and potentially of little value for policy or clinical decision-making¹⁷⁷. Despite the proliferation of value frameworks, ‘value’ remains an elusive target and a wider consensus about what dimensions of value to include may still be some way off, maybe with the exception of frameworks that are applied in clinical practice (as potential value dimensions are more restricted in nature).

The use of multiple criteria decision analysis (MCDA) methods has been proposed as an alternative methodological approach for assessing the value of health care interventions in different contexts, ranging from licensing decisions at the marketing authorization stage³³¹, to coverage decisions at the HTA stage¹⁹⁶, to treatment selection decisions at prescribing level³⁴⁰. MCDA methods can be used for quantifying benefits, risks and uncertainties in

order to aid the decision-making process, by considering an explicit set of criteria and their relative importance under a fully transparent process, while incorporating a wide range of stakeholder views to express a more societal perspective.

A methodological process towards the development of a robust MCDA framework observing key principles to ensure methodological robustness and consisting of different phases and stages for implementation in the context of HTA has already been proposed ³³². Possibly the most fundamental phase of the MCDA process with the highest impact on the overall outcomes relates to model building and the criteria selection phase, essentially directing the value concerns being addressed based on which the alternative options will be assessed against.

In this paper I focus on model building and criteria selection, describing the development of a generic value based model taking the form of a value tree for the purpose of assessing the value of new medicines in the context of HTA by capturing value for decision makers. Although in theory such a value tree can be generic for any type of health technology, including drugs, medical devices and other health interventions, for ease of illustration, I focus on new medicines. In Section 2, I discuss the theoretical foundations of MCDA, starting with the theoretical axioms of decision analysis that essentially address the question of “what is the basis of MCDA”. In section 3, I outline the methods for model building and the selection of different evaluation criteria. Section 4, presents the results involving the assembly of decision makers’ concerns into a comprehensive generic value tree, which provides insights into “what is value in the context of HTA”. Section 5 presents the discussion, introducing different MCDA methods and proposes ways to operationalise the value tree through a precise combination of MCDA techniques and addresses the question of “how to apply MCDA in HTA” through the use of a new value framework. Finally, section 6 draws the main conclusions.

5.2 Theoretical Foundations

Decision analysis was originally defined by Howard as “a logical procedure for the balancing of the factors that influence a decision” incorporating “uncertainties, values, and preferences in a basic structure that models the decision” ⁵². The logic behind decision analysis was described as divide and conquer whereby a complex problem is decomposed into simpler problems and each individual problem is analysed separately before all analyses are connected together, resulting in a program of action for the complex problem ⁵³. The methodology of decision analysis is simplified to 4 main steps, notably (a) structure the

decision problem; (b) assess possible impacts of each alternative; (c) determine preferences (values) of decision makers; and (d) evaluate and compare alternatives^{52,53}.

The starting point for this discussion is that the value of new medical technologies is multidimensional and not limited to their clinical effect or benefit. Besides the traditional dual clinical consideration of health benefits and risks, taking the form of efficacy and safety considerations, or the emerging dual economic consideration of health benefits and costs, taking the form of health outcomes and costs per outcome unit, other factors may also be important for determining the value of a new medicine. The severity and unmet need of the disease, the clinical novelty and convenience to patients, or the wider benefits to society have been at times perceived as important considerations of value to decision makers for the purpose of achieving efficient resource allocation³¹.

Based on that, the value of new medical technologies can be illustrated as a function of different evaluation parameters, as part of a linear additive model, namely:

$$Value = f(a_{ij}, b_{ij}, c_{ij}, d_{ij}, \dots, n_{ij}) \quad (5)$$

where a, b, c, d, \dots, n denote the different parameters of interest, i denotes the medical technology's value in regards to a particular parameter and j denotes the relative importance weight of the same parameter based on decision-maker or other stakeholder views. Additionally, questions remain on how to incorporate the views of all relevant stakeholders and how to derive relative weights for the different parameters.

Given that the value of new medical technologies is based on a multitude of value dimensions and the limitations of current approaches to value assessment, there is a need for an alternative methodological approach of value assessment that encompasses multiple value domains explicitly, therefore a decision analysis method that addresses multiple attributes of benefit is required.

Among methodological tools for assessing value quantitatively as part of decision-making process, MCDA could be indicated as an adequate method, ordering a set of alternative options based on the degree to which a number of different objectives are achieved^{1,168}. One of the main aims of MCDA methods is to enable decision makers reach a decision by facilitating them to learn and understand more about the problem, objectives, and values being faced, through organising and synthesising information of complex and conflicting nature². MCDA can facilitate decision making by explicitly integrating objective

measurement with value judgement while managing subjectivity in a transparent way, however it cannot act as a substitute to decision-making.

Having introduced the theoretical foundations of decision analysis and the overall MCDA process, the next step is to focus on the model-building phase by applying MCDA principles. As part of the model-building phase, the criteria selection stage is crucial, involving their identification and assembly into a hierarchical structure taking the form of a tree^{56,341}. The aim is to arrive at a generic value model for new medicines that can be adapted to capture all relevant dimensions of value across different decision-making contexts and therapeutic indications. Criteria represent the key concerns influencing a particular decision. Structuring all the criteria in the form of a tree is known as a *value tree* and provides an organized schematic representation of the various concerns under consideration by the decision-maker. The criteria-based evaluation of options is operationalised through the use of performance descriptors, either of a qualitative or quantitative nature, known as *attributes* which essentially measure the fulfilment of the criteria (Box 1).

Box 1: Definitions of decision analysis terminology

Criterion: an ‘individual measurable indicator’ of a key value dimension¹ or more precisely, a ‘particular perspective according to which alternative technologies may be compared’².

Attribute: a ‘quantitative or qualitative measure of performance associated with a particular criterion’², or in other words a descriptor of performance or impact requiring ordering of preference³

Value tree: an organized schematic representation of the various objectives, criteria and attributes under consideration.

5.3 Methods

5.3.1 Model Building Approaches

The three main steps in building a multi-criteria model are to (a) structure a value tree that identifies and represents the objectives or key concerns of decision makers, (b) define attributes at the bottom level of the value tree to measure the achievement of these objectives, and (c) select decision alternatives ²⁰⁴.

Structuring a value tree can generally be done using two approaches; either through a top-down approach (known as value-focused thinking), which is driven by the overall objective or value concern and is decomposed into lower levels of sub-objectives or sub-concerns; or a bottom-up approach (known as alternative-focused thinking), which is driven by the alternative options under consideration based on attributes that distinguish between them and which are grouped into higher levels of objectives and concerns ⁵⁴.

Given the aim of the study is to build a generic model that can subsequently be adapted and applied across different decision-making contexts, I used the top-down approach (value-focused thinking) ³, so that the model can reflect the overall value concerns of decision makers while being adaptable to different decision problems. In other words I aimed to incorporate the value dimensions of new medicines in general that decision makers want to capture as part of the evaluation process. With regards to the completion of the remaining two tasks (attribute definition and selection of decision alternatives), I recommend the adoption of the bottom-up approach (alternative-focused thinking) following the definition of the decision problem ³⁴², so that the model can become decision-specific to precisely assess the performance of alternative treatments as needed. As such, I suggest the completion of the two tasks and the model building process as part of particular applications, where the comparison of actual decision options feeds into the selection of precise attributes for distinguishing their value.

Overall, in order to complete the three main steps, I have proposed a “value-alternative hybrid thinking” ³³², under which the core structure of the value tree takes place as part of a top-down approach, and the definition of attributes is ultimately completed following the selection of the decision alternatives as part of a bottom-up approach.

Importantly, both criteria and attributes of the value tree must possess a number of properties for the results to be robust ³³²; these include being essential, non-overlapping and concise (for the case of criteria); unambiguous, comprehensive and direct (for the case of attributes); understandable, operational, and preference-independent (for both criteria and

attributes)^{2,204-206}. A discussion on these properties as well as the significance of adhering to them in the context of HTA to ensure a robust process for the application of MCDA is provided elsewhere.

5.3.2 HTA Adaptation Approach and Staging

In the context of HTA, it would be reasonable to assume that the value concerns (i.e. the criteria) of decision-makers would encompass all the key factors that are necessary for a comprehensive appraisal of value of a new medicine or health care intervention, with their relevant disease-technology *characteristics and impacts* on patients outcomes and health systems resources acting as descriptors of performance (i.e. attributes) for measuring the extent of satisfying these criteria. However, a universal or common set of decision-makers' value concerns (i.e. criteria) and their measures (i.e. attributes) are neither clearly defined, nor well-established, both within and across health systems. Therefore, a definition of *value* in the context of HTA is currently absent^{69,339}.

In order to capture different decision-makers' concerns in a comprehensive manner that ultimately would lead to the structuring of a generic value tree, a five-stage iterative model-building process was followed, involving both secondary and primary data collection and adopting a top-down 'value-focused thinking' approach, as shown in Figure 5.1. The five steps were as follows:

First, I conducted a systematic review of the peer-reviewed literature to identify what value dimensions are considered in the evaluation processes of HTA bodies in eight EU countries³³⁹. The Centre for Review and Dissemination guidance for undertaking systematic reviews in health care was followed²²⁶. The eight study countries (and their HTA bodies) were France (Haute Autorité de Santé, HAS), Germany (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, IQWiG), Sweden (Tandvårds- och läkemedelsförmånsverket, TLV), England (National Institute of Health and Care Excellence, NICE), Italy (Agenzia Italiana del Farmaco, AIFA), the Netherlands (Zorginstituut Nederland, ZIN (formerly College voor zorgverzekeringen, CVZ)), Poland (The Agency for Health Technology Assessment and Tariff System, AOTMiT) and Spain (Red de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) and the Inter-ministerial Committee for Pricing (ICP)). The rationale for their selection was the variation in their health system financing (tax-based vs. social insurance-based), the organisation of the health care system (central vs. regional organisation), the type

of HTA in place (predominantly economic evaluation vs. predominantly clinical benefit assessment) and the perspective used in HTA (health system vs societal), so that the sample is representative of different health systems and HTA approaches across Europe. Inclusion criteria for the review were the following: (i) English language, (ii) evidence from the eight study countries of interest (and their respective agencies), (iii) HTA context from a national coverage perspective, and (iv) a publication date between January 2000 to January 2014¹⁰⁶. The electronic databases of Medline and Social Science Citation Index were searched for English peer review literature using the keywords “health technology assessment” OR “value assessment” AND “pharmaceuticals” OR “methodologies”. Reference lists of the selected studies were screened, and the HTA bodies’ websites were searched for any published guidelines. Material on the “type of evidence and evaluation criteria considered in HTAs” was collected and analysed, along with material on the “responsibilities and structure of national HTA models and processes”, “methods and techniques applied in HTA”, and “outcomes and implementation of HTAs”. The identified value dimensions sourced from the “evidence and evaluation criteria” component of the review would form the fundamental domains of the model, taking the form of top-level criteria groups (i.e. clusters) and informing their decomposition into lower level criteria, which comprised the core structure of the value tree.

Second, literature findings were supplemented with expert consultation, where national agency HTA experts from the countries of interest were invited to review and validate the results. This took place because upon an early communication of the preliminary results with the partners of the Advance-HTA project consortium²²², it became obvious that in a few cases the evidence from the peer review literature may have been outdated and, in some cases contradictory, and did not reflect actual practices. As a result, the findings of the systematic literature review were validated with national experts from the agencies in question.

Third, I incorporated findings from other relevant literature of health care interventions evaluation, including grey literature, to identify value concerns of decision makers that might not be reflected as part of the current or formal HTA evaluation criteria in place. I considered studies on the benefit-risk assessment of new drugs from a licensing perspective^{73,331,343-350}, value based pricing and assessment from a payer perspective

¹⁰⁶ Product-specific technology appraisals or evaluation studies and conference proceedings or records with no abstract available were excluded.

31,139,154,190,351-354³⁵⁴, and patient access from a social responsibility perspective^{176,355}. The findings from this step supplemented the lower level criteria and their decomposition into bottom-level sub-criteria or attributes.

Fourth, following the completion of the above steps, the emerging structure of the value tree and its criteria were subjected to a detailed consultation with 28 HTA experts who provided feedback on the comprehensiveness of the model, but also on its perceived usefulness and practical limitations, as part of the Advance-HTA project²²². These experts were selected because of their relevant expertise and because they acted as partners, Scientific Advisory Board members and affiliated scientists in the Advance-HTA project and included health care professionals (e.g. clinicians, nurses, pharmacologists), methodology experts (e.g. health economists, HTA experts, statisticians) patient representatives and policy-makers/regulators, who were affiliated with a wide range of academic and research institutions at international level, including nine academic-research institutions with health economics and/or HTA centres¹⁰⁷, four HTA bodies¹⁰⁸, one HTA research network¹⁰⁹, one coordinating patient and health care professional organisation¹¹⁰ and one international public health organisation¹¹¹. Based on the feedback received from these experts, the structure of the value tree was revised in an iterative manner, mainly informing the bottom-level sub-criteria or attributes.

Finally, following a series of dissemination activities as part of the Advance-HTA project involving presentations and seminars as part of capacity building workshops, feedback was collected from a wide range of stakeholders (mainly decision-makers from ministries of health, health insurance organisations and HTA agencies) and key opinion leaders across settings beyond the eight countries identified originally, where the aim was to validate results and further enhance the encompassing nature of the model by capturing additional expert and wider geographical perspectives³⁵⁶⁻³⁵⁹. Specifically, the value tree was disseminated at four workshops attended by a total of 230 participants from the Latin

¹⁰⁷ London School of Economics and Political Science (LSE), London School of Hygiene and Tropical Medicine (LSHTM), Instituto Superiore di Sanita (ISS), Universidad de Castilla – La Mancha (UCLM), Institute za Ekonomska Raziskovanja (IER), Technische Universitaet Berlin (TUB), Escuela Andaluza de Salud Publica (EASP), Universite Paris XII - Val de Marne (UPEC), and University College London (UCL).

¹⁰⁸ NICE International, Agencja Oceny Technologii Medycznych (AOTM), Tandvårds- och läkemedelsförmånsverket (TLV), and Haute Autorité de Santé (HAS).

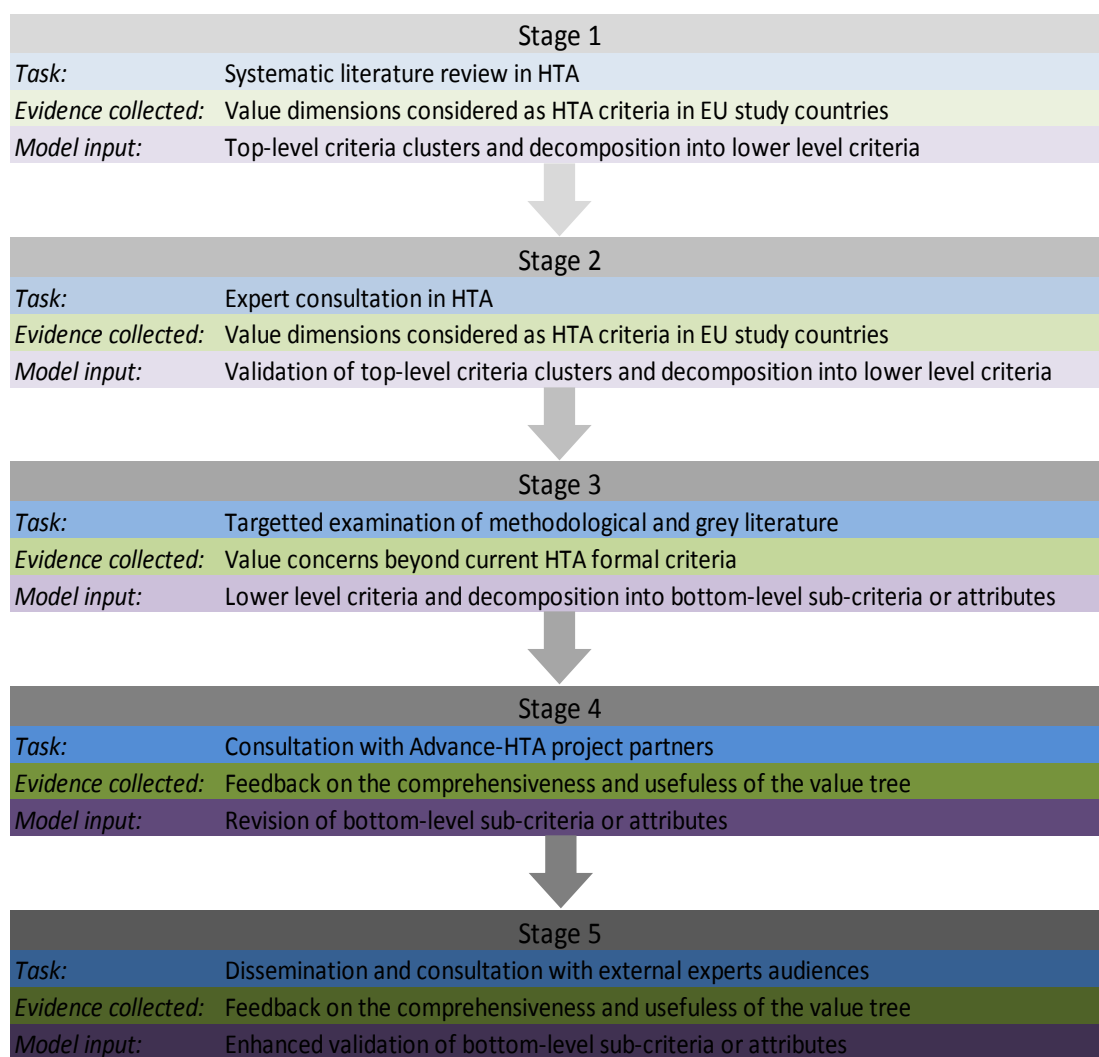
¹⁰⁹ European Network for Health Technology Assessment (EUnetHTA).

¹¹⁰ European Brain Council (EBC).

¹¹¹ Pan American Health Organisation (PAHO).

American (November 2014 and September 2015) and Eastern European regions (September 2014 and 2015), in order to capture perspectives from low- and middle-income countries in these regions in the form of smaller focus groups of 10-15 participants.

Figure 5.1: The five-stage criteria selection process for structuring the value tree as part of a top-down ‘value-focused thinking’ approach.



Throughout the five stages of the model-building process, the various pieces of evidence collected to inform criteria selection and their structuring into a value tree took place in alignment with decision theory principles, aiming to satisfy the required criteria properties so that the model produced is rigorous and the analysis outcome robust²⁰⁴. For example, an additional criterion would be added if a value concern was not captured in the initial set of criteria so that all the essential value concerns of decision makers could be

addressed. Further, if a particular criterion was perceived to reflect the same value concern as another criterion, one of them was removed in order to avoid double-counting. Similarly, two individual criteria could be replaced if their underlying concern could be reflected from a single criterion on the basis of conciseness; generally, only the smallest number of criteria required for evaluation should be included, in order to strive for simplicity. Individual criteria could be replaced with other criteria if their meaning was not clear in order to improve comprehension; similarly, individual criteria would be replaced if their measurement was not possible, in order to ensure high levels of functionality. Finally, if by assessing the value of one criterion it became evident that knowledge on the performance of another criterion would be required, then the two criteria would be aggregated into a single one in order to become preference-independent.

Despite this cautionary approach, the value tree aims to capture a comprehensive generic set of value concerns that can be adapted to different decision-making contexts, problems, indications or treatments. As a result, the individual criteria could be compared across different decision-making settings and jurisdictions, to reflect similarities and differences of decision-makers' value concerns as evidenced by the inclusion of particular criteria and the way of their measurement (i.e. precise attributes). As a result, following the completion of the model-building phase involving the precise attribute definition and the selection of alternative treatment options for the case of specific decision problems, some of these criteria might not satisfy the required properties in which case the underlying issues should be addressed with caution. For example, following the operationalisation of two criteria with specific attributes it might become evident that there is possible double-counting between them, in which case one of the two would have to be excluded. A few cases in which the theoretical properties of some criteria could be put into question have emerged following consultation with experts; these are discussed in the respective sections. Overall, the five-stage process took place between February 2013 and end-October 2015.

5.4 Results – The Advance Value Tree

The findings from the systematic literature review and expert consultation in HTA which acted as the first and second stages of the model-building process are briefly presented, mainly because it informed the primary identification of value dimensions which established the core structure of the value tree. Then the completed value tree is outlined and the logic behind its various components and value dimensions discussed, including the overall criteria grouping and decomposition from top-level criteria clusters into lower level criteria and

bottom level sub-criteria or attributes. Finally, I briefly suggest how to deal with quality of evidence, and more precisely with clinical validity concerns.

5.4.1 Primary Identification of Value Dimensions: Findings from the Systematic Literature Review and Expert Consultation in HTA

In total, 2778 articles abstracts were screened out of which 255 articles were selected to be read in full based on their relevance, with the content of 101 articles being ultimately used. The main groups of value dimensions that were found to be considered as evaluation criteria among the study group of European countries as identified through the first and second stages of the model-building process included: (a) burden of disease, (b) therapeutic impact, (c) safety profile, (d) innovation level, and (e) socioeconomic impact³³⁹. Individual value dimensions falling under these main groups of evidence, together with their intensity of use by each country are shown in Table 6.1. Based on the available evidence, these five clusters of evaluation criteria were perceived to comprise the minimum critical aspects of value dimensions of interest to decision-makers for evaluating the value of new medicines as part of HTA, providing the core foundation of the value tree.

5.4.2 Incorporating the Value Dimensions into a Generic Model: The Advance Value Tree

Ultimately, the resulting generic value tree spans three levels of evaluation criteria (top-, middle- and bottom-level), where top-level groups of criteria (i.e. criteria clusters) are decomposed into middle-level criteria and bottom-level sub-criteria or attributes, relating to the five key value domains described above that can be explicitly measured and assessed: (a) burden of disease (BoD), (b) therapeutic impact (THE), (c) safety profile (SAF) (d) innovation level (INN), and (e) socioeconomic impact (SOC). With the exception of BoD cluster, which relates to the disease or indication of interest, the remaining four clusters relate to the impact or characteristics of the medicine. The hierarchical representation of the three levels of evaluation criteria forms the different components of the value tree, which I called ‘Advance Value Tree’, as shown in Figure 5.2. In particular evaluation contexts, the hierarchical organisation of these value concerns could be perceived to represent a combination of objectives and indicators rather than a value tree in their strict sense.

Although additional types of value concerns might exist falling under other categories^{66,360,361}, such as efficiency (e.g. cost-effectiveness), equity (e.g. priorities, fairness, ethics, etc.) and implementation complexities (e.g. organisational, skill and

legislative requirements), they are not included as criteria because they would contradict with the desired criteria properties and the adopted scope of “value”. Strictly, these dimensions do not represent intrinsic disease-technology characteristics or impacts on patient outcomes and health system resources but, instead, go beyond to capture extrinsic characteristics and impacts that generally depend on the value tree’s core variables, or other features of the health care systems reflecting a wider health systems goals and building block perspective ³⁶².

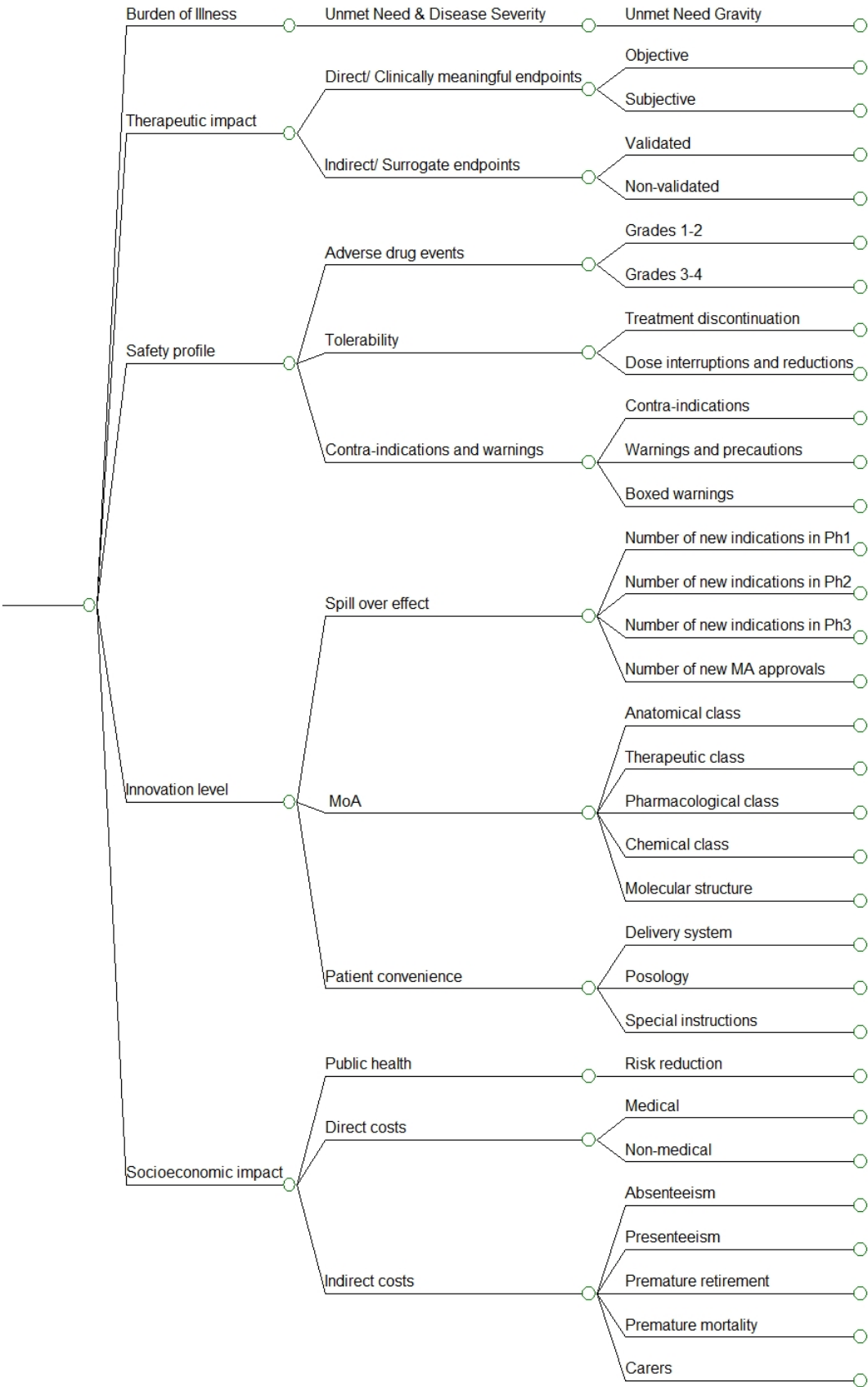
For example, efficiency is a composite concept comprising two components, i.e. cost and benefit, with the latter already reflected in the model and thus its inclusion would violate the principle of non-overlap leading to double-counting. Concerns relating to equity, implementation complexities, and other characteristics of the overall health systems’ context (e.g. stakeholder pressure and political power), are usually of subjective nature and not easily quantifiable, therefore making it hard to operationalise. All these extrinsic value dimensions do not relate to the “value” of a new medicine *per se*, but instead depend on the settings of the particular health system under consideration. They could therefore be considered or incorporated on an optional basis and as needed for the particular decision context and problem in question, possibly through the use of other analytical frameworks ^{90,363}.

Examples of iterations in the value tree that have resulted from consultations with experts include the aggregation of ‘safety’ and ‘tolerability’ criteria into a single ‘safety & tolerability’ criterion when a potential overlap between their measures became evident following discussions with clinicians (described in the ‘Safety profile’ section below), or the addition of a ‘Carer’ sub-criterion under the ‘Indirect costs’ criterion to capture the wider socioeconomic impact of a treatment following discussion with patients.

Table 5.1: Value dimensions considered as evaluation criteria among the study group of European countries and their intensity of use

[illegible]

Figure 5.2: The generic set of value dimensions for new medicines Advance Value Tree



5.4.2.1 Burden of disease

‘Burden of Disease’ (BoD) forms a special set of value dimensions as they do not relate to the medical technology itself, but to the disease it is indicated for and, as such, could encompass the severity and unmet need of the disease the treatment addresses^{352,364}. Severity of the disease relates to the condition’s degree of seriousness in respect to mortality and morbidity-derived disability, which could be defined on the basis of disability-adjusted life years (DALYs) lost³³⁸, or the expected remaining life years adjusted for their quality of life³⁶⁵. Unmet need reflects the availability of treatments, essentially the degree to which there are existing treatments³¹ and could relate to methods of diagnosis, prevention or treatment³⁶⁶. Because these dimensions are preference-dependent (i.e. in order to assess the impact of unmet need one needs to consider the severity of the disease) they are operationalised through the use of a single aggregated attribute which could be defined as the gap between the health status that patients with a particular medical condition can attain using existing¹¹² medical interventions and the health status they could expect if they did not have that medical condition, or in other words “the number of QALYs lost by a patient because of their condition”³⁵². Therefore, the BoD attribute reflects the difference in the years of life remaining (adjusted) with the respective HRQoL for patients receiving existing technologies, versus the years of life remaining (adjusted) with the respective HRQoL for healthy individuals (of the same age). The larger the difference or the gap between the two (i.e. diseased vs. healthy states), the higher the disease burden.

The size of the population being affected by the disease, and the possible differentiation of a disease on the grounds of high prevalence or low prevalence is not taken into account because it could be perceived as unethical. The justification of a special status of a disease based on its prevalence would be questionable, as it entails valuing one disease differently to another because they are a more common vs. a less common disorder³⁶⁷. The only justification for providing special status to rare diseases would be on equity grounds, using the rationale that “patients suffering from rare conditions should be entitled to the same quality of treatments as other patients”³⁶⁸. However, this concern is already addressed through the unmet need criterion and the associated (un-) availability of effective treatments; consequently, the inclusion of a criterion for capturing ‘population size’ would essentially

¹¹² Ideally the “existing” medical intervention or treatment should not be one of the options being assessed so that there is no double counting with the Therapeutic Impact cluster. “Existing” could be defined at the HTA level, i.e. in respect to what has been approved for reimbursement, and if nothing has been approved then in terms of best supportive care (BSC).

lead to double-counting and, therefore, is not inserted. Valuing a condition ‘more’ or ‘less’ as a result of prevalence would be incompatible with other equity principles and theories of justice ³⁶⁷.

5.4.2.2 Therapeutic impact

For most medical conditions there is a set of multidimensional health outcomes that need to be jointly used to capture overall patient benefit including survival, functional status, sustainability of recovery and others, including complications ³⁶⁹. Outcomes are distinct from biologic indicators, the former usually relating directly to health status in contrast to the latter which act as predictor of results.

The ‘Therapeutic impact’ cluster captures clinical benefit by measuring both direct outcomes and indirect indicators relating to the efficacy and/or the effectiveness of an intervention with the view to reflecting the health status but also disease recovery, progression or prevention (complication outcomes are considered separately in the “Safety Profile” cluster discussed in the next section).

In order to distinguish between direct outcomes and indirect indicators, based both on the literature and subsequent expert consultation, “Therapeutic Impact” criteria have been divided into (a) direct, clinically meaningful outcomes and (b) indirect, surrogate indicators respectively; the latter are used as substitutes for direct endpoints being usually disease-specific in nature (e.g. HbA1c for complications in diabetes mellitus, prostate-specific antigen (PSA) for prostate cancer, blood pressure for cardiovascular disease). Specifically, a direct endpoint is defined as a “characteristic or variable that reflects how a patient feels, functions, or survives” while an indirect endpoint is defined as a “biomarker intended to substitute for a clinical endpoint” which is “expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence” ³⁷⁰ (page 91). In turn, the definition of a biomarker states that it is a “characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” ³⁷⁰(page 91).

However, both types of endpoints are useful and, for specific indications, a biomarker (i.e. indirect indicator) might be more directly related to real health status than self-reported outcomes (i.e. direct outcome); this could be a reason why an increasing number of studies in health economics use biomarkers instead of self-assessed measures of

health. Below I propose a further decomposition of direct meaningful endpoints and indirect surrogate endpoints, once again guided by the five-stage evidence collection process.

Direct meaningful endpoints

Direct clinically meaningful endpoints can be divided into objective and subjective endpoints ³⁷¹ and, possibly, other health related outcomes ³⁷². Objective direct endpoints mainly refer to survival, disease exacerbation/alleviation and clinical events. The availability of evidence on different endpoints depends on the study designs adopted by the respective clinical trials for evaluating the clinical benefit of interventions and, therefore, would depend on their type or nature.

Subjective direct endpoints mainly relate to HRQoL and disease symptoms. HRQoL embraces the broader concept of *health* that includes physical, emotional and social wellbeing ³⁷³, by including both personal health status and social wellbeing ³⁷⁴ usually being “subjectively” assessed through patient reported outcome measures (PROMs). HRQoL is multidimensional ^{375,376} and can generally be measured through the use of generic PROM that provide a summary of HRQoL attributes through the production of health utilities ³⁷⁴, such as SF-36 and EQ-5D instruments¹¹³. Generic PROM instruments do not target specific population groups and for the case of particular disease states they might not be sensitive enough to adequately capture the impact across all HRQoL dimensions in which case disease specific instruments might be needed. From a societal perspective all these dimensions should be considered both for the cases of patients and carers.

Importantly, following consultation with experts, it became apparent that the assessment of particular combinations of objective and subjective endpoints as for example OS and HRQoL (e.g. through EQ-5D) might be preference dependent. For example, in the context of a metastatic cancer setting, stakeholders highlighted that their preferences related to the performance of different therapies in terms of OS would be meaningful only if the HRQoL performance was also known for the same therapies. Where preference dependence is evident, the dependent attributes should be combined into a single attribute, as for example QALYs, essentially aggregating them into a common attribute capturing both OS and HRQoL considerations.

¹¹³ Generic PROM instruments do not target specific population groups and for the case of particular disease states they might not be sensitive to adequately capture the impact across all HRQoL dimensions in which case disease specific instruments might be needed.

Indirect surrogate endpoints

Indirect surrogate endpoints can be divided into validated and non-validated. Validation of a surrogate endpoint is the process of retrospectively linking it to the actual clinical endpoint (or outcome), i.e. demonstrating a relationship between the two by evaluating how well the surrogate endpoint predicts the clinical outcome of interest ³⁷⁰. Besides the existence of a strong statistical correlation between the surrogate and the clinical endpoint that is needed in order to entail an accurate prognosis of the clinical outcome, i.e. ‘individual-level surrogacy’, there is also a prerequisite for clinical correlation through the existence of biologic plausibility, i.e. scientific evidence on the causality between the disease, surrogate and outcome; importantly, there is also a need for demonstrating “trial-level surrogacy” which refers to the correlation between a change in the surrogate and a change in outcome due to a therapeutic intervention ³⁷⁷. Although trial-level data are usually coming from multiple trials or units for the same type of intervention, for a surrogate to be validated it means that its “validity is generalizable to include other interventions that affect the surrogate endpoint” ³⁷⁰(page 93). As a result of these requirements, surrogate endpoints are rarely validated. Generally, most only manage to predict the clinical endpoint or outcome of interest whilst others fail to predict the clinical endpoint and are used only as a measure of biological activity ³⁷⁸.

5.4.2.3 Safety profile

The safety profile of an intervention comprises information relating to the degree of its safety and toxicity for the indicated patient population of interest. It is usually measured through safety and tolerability, but can also be reflected through any contra-indications for its use and special warnings (and precautions) for any particular sub-populations.

Safety and tolerability

Safety has been traditionally reflected through the incidence of adverse events. An adverse event (AE) refers to an adverse outcome occurring when a drug is administered to a patient or at some time afterwards, in which case the drug might or might not be the cause of the AE ³⁷⁹. If such outcomes can be attributed with some degree of probability and through a causative link to an action of the drug, then they are known as adverse drug events (ADEs).

The magnitude of both AEs and ADEs is measured through the combination of their seriousness and their frequency (or probability of taking place). Given that seriousness and probability of adverse events seem to be preference dependent, an aggregated attribute

reflecting ‘seriousness and frequency’ would be needed. Seriousness could be operationalised through the use of the Common Terminology Criteria for Adverse Events (CTCAE) classification, which contains five grades of adverse events ranging from mild (Grade 1) to death (Grade 5), with three grades in between (moderate, severe, life-threatening)³⁸⁰. However, for cases that overall survival is already incorporated in the value tree under the therapeutic impact cluster, then considering Grade 5 adverse events which relate to deaths would constitute double counting and should be excluded. In turn, given that the distinction of adverse events between mild and moderate and between severe and life-threatening is often subjective, they could be aggregated and categorised as “non-serious” versus “serious”, the former comprising Grade 1 and 2, and the latter comprising the Grade 3 and 4 adverse events. Frequency could be operationalised through the use of absolute incidences (percentages), or, alternatively, using a frequency classification system such as the Naranjo scale which expresses frequency in terms of definite, probable, possible or doubtful³⁸¹.

Tolerability refers to the overall ability of the patient to tolerate the intervention, mainly in regards to bearing and enduring any adverse events. It is usually reflected through the variables of treatment discontinuation and treatment interruption or reduction, measured either as the proportion of patients discontinuing the treatment (or interrupting/reducing its dosing) or the time to treatment discontinuation (TTTD) from the treatment.

Importantly however, as it became apparent following consultation with clinical experts, in the case of discontinuation (or interruption or reduction) of a treatment being due to the incidence of known ADEs, then incorporation of both measure types could lead to double-counting and therefore caution would be needed to choose the most appropriate between the two.

Contraindications, special warnings and precautions

Contra-indications refer to factors¹¹⁴ that act as a reason for an intervention not to be used by a patient, thus having an impact on the number of potential patients using and benefiting from it. Strictly speaking, contraindications can be categorised as absolute and relative. In the former there are no circumstances under which the patient might use the intervention, compared to the latter whose risk might be outweighed in favour of other considerations (e.g. x-rays for a pregnant woman with a risk of having a child with birth defects). Special

¹¹⁴ usually related to its risk and safety but can also be related to its benefit and efficacy/effectiveness

warnings and precautions for use are designed with the view to notifying potential risks associated with the use of the intervention in regards to specific patient sub-populations with particular characteristics. These characteristics mainly relate to the administration of concomitant medication, the coexistence of other accompanying diseases and the presence of idiosyncratic patient pathological features that as a result might influence the expected action of the drug; as a consequence, caution in the form of careful monitoring is usually suggested.

In the real world, some patients with known contra-indications (or special warnings and precautions) for a treatment might end up receiving it, in which case incorporation of any evident ADEs or tolerability consequences that can be related to them as attributes in the value tree, would lead to double counting between the two and should therefore be addressed with caution, similar to the case between ADEs and discontinuation.

5.4.2.4 Innovation level

“Innovation” in the context of medical technologies, ranging from biopharmaceuticals and diagnostics to medical devices, is a complicated concept lacking a universal consensus ³⁸². From a patient perspective, “innovation” mainly relates to “therapeutic innovation” requiring *novelty of effectiveness*: value needs to be created by generating improved health outcomes that were previously unattainable, the degree of which could be assessed through the combination of the significance of the unmet medical need the drug addresses and the extent to which it improves the health outcomes for that need ³⁸³. Below I discuss secondary innovation dimensions over and above “therapeutic innovation”, given that the significance of unmet medical need and the extent of health outcomes improvement are captured under the “Burden of Disease” and “Therapeutic Impact” clusters respectively. These dimensions include (a) the mechanism of action, (b) the technology’s spillover effects, and (c) patient usefulness (convenience).

Mechanism of action

The innovativeness of new medicines can be differentiated according to their type or nature, based on whether they offer a novel mechanism of action, or whether they act through more typical mechanisms of action. A relatively practical and objective way of measuring that would be through the use of WHO’s Anatomical Therapeutic Chemical (ATC) Classification

System¹¹⁵. Moving from general to specific, this classification has 5 main dimensions relating to anatomical, therapeutic, pharmacological, chemical and molecular levels which could act as the respective criteria. By this logic, a medicine characterised by a novel therapeutic action, would be more innovative than a medicine with a novel pharmacological action and so on. In other words, the broader the level at which the drug (or combination) differentiates as ‘original’ compared to current existing alternatives, the more innovative the nature of that drug (or combination) would be.

As a result, the drug’s relative market entrance in regards to the different innovation subgroups (i.e. ATC levels) could act as the respective attributes, in order to reflect, for example, whether it is first-in-chemical-class (i.e. first entrance of a technology at level 4, chemical subgroup level), second-in-pharmacological-class (i.e. second entrance of a technology at level 3, pharmacological subgroup), and so on.

Spill-over (dynamic efficiency)

Any type of innovation can have R&D spill-over effects that can lead to the development of subsequent innovation(s), entailing a certain degree of diffusion of scientific knowledge and/or technical know-how. Innovation ‘spill-over effects’ could be defined as “the R&D positive externalities that can lead to the development of subsequent innovation(s)”, thus essentially relating to dynamic efficiency in regards to long-term product innovation at future market conditions. These effects could take the form of internal (within the innovator) or external (outside the innovator) effects³⁸⁴. As Lipsey and Carlow have argued, ‘major radical innovations never bring new technologies into the world in a fully developed form’, but “appear in a crude and embryonic state with only a few specific uses”³⁸⁵. Instead, successive improvements are accumulated through the processes of ‘learning by doing’ and ‘learning by using’³⁸⁶, the former referring to the improvement of workers’ skills at the manufacturing level³⁸⁷, while the latter relating to enhancements of knowledge at the level of utilisation by the final user³⁸⁸. Subsequent to the market entry of a new drug, new uses (i.e. new indications) for the same drug could be uncovered following its investigation for

¹¹⁵ In the Anatomical Therapeutic Chemical (ATC) classification system, the active substances are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Drugs are classified in groups at five different levels. The drugs are divided into fourteen main groups (1st level), with pharmacological/therapeutic subgroups (2nd level). The 3rd and 4th levels are chemical/pharmacological/therapeutic subgroups and the 5th level is the chemical substance. The 2nd, 3rd and 4th levels are often used to identify pharmacological subgroups when that is considered more appropriate than therapeutic or chemical subgroups.

use by patients with other diseases, or next generation drugs could be successively developed either for the same disease indication or a different one. New scientific knowledge and/or technical know-how could be diffused into other contexts/sectors, leading to innovations therein. Because the impact on future innovation is uncertain, it would be practically challenging if not impossible to be predicted and assessed at a single point in time³⁸⁹, as for example the rate at which performance improvements take place and the speed at which new uses are discovered. In reality, a significant portion of any such spill-over effects will need time to materialize requiring them to be in the market for a period of time.

Although these effects can relate to new uses for the same technology, new technologies for the same use, or new technologies for new uses, only the first ones are considered here because the latter two are not operational¹¹⁶. Essentially, this criterion would refer to the extent to which the drug has a spill-over effect in the context of expansion into new indications and could be operationalized by examining the number of new indications for which the drug is investigated for at each stage of clinical development (e.g. Phase I, Phase II, Phase III, Marketing Authorisation phase).

Patient usefulness (ease and convenience)

Aspects relating to patient usefulness would be another group of innovation-related dimensions. Satisfaction of patients with medical care has been shown to correlate with compliance, i.e. the willingness and ability to follow health-related advice including adherence to prescribed medication, while also acting as predictor of future compliance³⁹⁰⁻³⁹². Satisfaction can be categorised into a range of different domains, including symptom relief/efficacy, side effects, ease and convenience, impact on HRQL, general satisfaction, and other domains specific to the given question³⁹³. Given that most of these domains are associated with the health outcomes of the treatment, or are of a generic nature, only ease and convenience are considered here, essentially relating to mode of administration, dosing schedule, medication restrictions, and product-specific designs. Poor compliance has been reported as the most common cause of non-response to medication, with evidence supporting

¹¹⁶ It would be very hard, if not impossible, to identify whether a new technology has been developed solely from the R&D process of another technology, hence it is hard to establish a causative link. For example, for the case of a new technology developed by the same manufacturer of the parent technology, it could be the case that actually the later technology was the one that helped with the development of the first technology, but delays in the manufacturing and/or regulatory processes caused one to be marketed before the other. For the case of a new technology developed by a different manufacturer, it could be the case again that the second manufacturer actually had developed on its own (from in-house R&D) the new technology.

that better health outcomes are produced when patients adhere to treatment recommendations compared to those who do not adhere ³⁹⁴.

Therefore, any improvements on the above can be translated into greater patient satisfaction and possibly better treatment compliance, which could lead to improved health outcomes, either through an increase in clinical effectiveness or due to a decrease in adverse events. In turn, individualising treatment and minimising its complexity has been proposed for encouraging adherent behaviour ³⁹⁵; for example, once-daily dosing had been suggested as an important part of enhancing compliance, patient convenience and regimen simplification in hypertension ³⁹⁶, among others. Given that patient health outcomes are already incorporated in the value tree, it would be expected that it is necessary to assume an explicit disconnection between better patient satisfaction and improved health outcomes in order to avoid double-counting. This theoretical rationality would support the notion that greater satisfaction through improved ease and convenience is a critical value dimension of new medicines in its own right, contributing to product novelty.¹¹⁷

Specific aspects of improved ease and convenience can include less invasive delivery systems, improved posology because of reduced dosing frequency or duration of administration and abolition of special administration instructions (e.g. no instructions vs. “take without food”). An altered delivery system could involve a change in dosage form (e.g. tablet, liquid, spray, gel) or a change in route of administration (RoA) (e.g. oral, subcutaneous, transdermal), but because these two variables are almost always correlated (e.g. a tablet will most probably administered orally, and a gel will most probably administered transdermally), both aspects could be captured through a common ‘delivery system’ attribute reflecting the combination of dosage form and RoA. Differences in posology could be captured by an attribute reflecting the combination of dosing frequency in a given time period and treatment duration of each dose. However, following consultation with patients it became apparent that in order to assess a treatment’s ‘posology’, one might need to consider the treatment’s ‘delivery system’, therefore suggesting for preference dependence between the two. As a result, the two types of value concern were aggregated into a common ‘delivery system and posology’ criterion. Special instructions could be assessed by an attribute reflecting the existence of any guidelines accompanying the administration of the treatment (e.g. with vs without food, crushing tablets, etc.).

¹¹⁷ Nevertheless, in reality a strict dissociation between satisfaction and health outcomes might not be required in the first place because of the gap between efficacy data and effectiveness data (see Appendix to Chapter 5).

Alternatively, the medication regimen complexity index¹¹⁸ (MRCI) could be used as an all-inclusive proxy attribute to operationalise the assessment of ease and convenience as it incorporates a number of medication regimen aspects, including dosage forms, dosage frequency and administration instructions³⁹⁷.

5.4.2.5 Socioeconomic impact

Socioeconomic impact dimensions are used to incorporate any other concerns or benefits in the wider context, which mainly relate to (a) public health and (b) economic considerations.

Public health impact

Public health impact is mainly associated with any risk reduction in transmitting and developing the disease under consideration or any other disease within the broader population, thus reflecting a societal dimension of prevention. The levels of risk reduction could range between no risk reduction, to reduction of prevalence risk factors, to reduction in transmission, to prevention and prophylaxis from the disease⁹⁰, and as a result such a variable would be more applicable to the case of infectious diseases and diseases with known risk factors. The emergence and dissemination of resistance among pathogenic bacteria to the available antibiotics used in medical practice, leading to the latter's drop in effectiveness would be a good example. A hypothetical new class of antibiotics that can effectively treat a bacterial strain known to have developed antimicrobial resistance, would offer an important risk reduction of the disease within the broader population by 'dealing with resistance' and inhibiting its transmission. The added value of such a drug would materialise through the combination of improved health outcomes (as reflected via the "Therapeutic impact" cluster) and through a risk reduction in disease transmission as reflected via the "Public health impact" dimension in the socio-economic impact cluster.

Economic impact

Economic impact dimensions reflect the economic burden of the disease and can be mainly divided into direct costs and indirect costs. Cost of illness studies are used to identify and measure all costs related to a particular disease and although different studies can employ different methodologies and designs they usually adopt a common classification of cost

¹¹⁸ An index for medication administration complexity, being adjusted to the dimensions of dosage form, dosing frequency and additional administrative directions needed.

types ³⁹⁸. Direct costs can be either medical or non-medical; medical costs are the costs resulting directly from the disease treatment and include diagnostic tests, prescription drugs, inpatient care (hospital or physician), outpatient care (physician or ER), nursing home care, rehabilitation care and home health care including any disposable or replacing items (e.g. prosthetic limbs). Direct non-medical costs refer to non-healthcare costs and include costs related to transportation, relocation, household, comfort/rehabilitation items, property alterations and counselling services.

However, including ‘cost’ as a value dimension is prone to criticism, mainly because criteria should be conceived as attributes of benefit ²⁰⁷. For that reason ‘impact on costs’ rather than absolute costs are considered, looking at the marginal difference versus an alternative option that could act as a neutral benchmark, being exclusive of the purchasing costs of the drug; this comparator could fall outside the scope of the analysis and for example could be best supportive care of the case of an oncology indication. Options’ purchasing costs could be incorporated in the analysis later on to establish the efficiency of the technology by considering its total value in relation to its total cost.

Indirect costs reflect productivity losses arising from patient absenteeism, presenteeism, premature retirement, and premature mortality ³⁹⁹. In addition, as it became evident following consultation with patients, indirect costs for carers (i.e. caregivers) are an important dimension of productivity losses that should be considered as part of a societal context and relate either to financial costs (in case they are not already included in the medical and non-medical costs) or time-off work.

5.4.3 Dealing with Quality of Clinical Evidence

In the context of health care decision-making, several attributes of evidence quality could be identified including adherence to the requirements of the regulator or decision-maker, completeness of reporting according to the regulator’s guidelines, consistency of reporting with the sources cited, relevance of evidence to the context in question and validity of evidence in regards to scientific standards or methodological guidelines of research ⁹⁰.

Focusing on the validity of clinical evidence, which would be mainly relevant for evidence feeding the performance measurement of alternative options under the therapeutic impact (THE) cluster, this could be decomposed into internal and external validity characteristics. Internal validity refers to the extent to which an observed effect can be attributed to the true effect of the intervention under investigation, or, in other words, the extent to which it is free

from bias, in contrast to external validity which relates to whether the clinical effects found in a clinical study (e.g. RCT) can also occur outside of the study's settings, i.e. beyond experimental settings in real clinical settings^{226,400,401}. Rather than operationalising these evidence quality (i.e. validity) concerns through their incorporation as criteria and attributes in the value tree, 'penalty' functions could instead be applied when the clinical evidence used for the alternative options are of different quality. This is because it would not be methodologically robust to incorporate quality of evidence concerns as criteria; doing so would entail that the quality of evidence used to assess the performance of the options across the respective criteria could be compensated by the options' performances across the same criteria.

A penalty factor could be applied through a function that multiplies it with the performance of the alternative options across the relevant THE criteria or their respective value scores, in order to adjust them for their validity. This penalty factor would be tailor-made for the particular decision-making context, being defined based on the relative importance of evidence validity as a source of concern to decision-makers. For example, assuming that a weak internal validity is associated with the comparative treatment effect of a new drug due to inadequate allocation concealment and lack of double-blinding, a "strict" decision-maker might be willing to "discount" the clinical performance of the drug on some of its clinical endpoints up to X% of their original level. The lower the validity of the evidence, the higher the impact that the penalty functions would have on the performances of the drugs or their respective value scores across the relevant THE criteria. However, it should be noted that the concept of penalty function should generally be used with caution because penalty functions may be incompatible with the use of an additive value model (unless they are used within the attributes).

A useful categorisation for the quality of clinical evidence is provided by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, a framework developed for the purpose of producing consistent clinical guidelines in terms of rating quality of evidence and grading strength of recommendations⁴⁰². According to GRADE, *high quality* evidence could be defined as evidence for which 'further research is very unlikely to change our confidence in the estimate of effect', in contrast to *very low quality* evidence for which 'any estimate of effect is very uncertain'⁴⁰³. In turn, the Cochrane Collaboration tool for risk of bias and the RE-AIM framework could be applied for the assessment of internal and external validity respectively, the findings of which could then be used for the estimation of the penalty factor by feeding the penalty function.

Based on a systematic classification of internal validity sources, the Cochrane Collaboration has developed a tool for assessing the risk of bias in RCTs, for a series of items covering different bias domains and their sources ⁴⁰⁴. These bias domains are broken down into seven different types based on which RCTs can be assessed and rated (as “low”, “unclear” or “high” risk of bias), namely *selection bias* (with sources of bias including random sequence generation, and allocation concealment), *performance bias* (blinding of participants and personnel), *detection bias* (blinding of outcome assessment), *attrition bias* (incomplete outcome data), *reporting bias* (selective reporting) and *other bias* (i.e. anything else).

On the other hand, the RE-AIM framework which was originally developed by Glasgow and colleagues to evaluate the public impact of health interventions, could be partially applied in order to assess the external validity of clinical trials, mainly through the dimensions of reach, adoption, implementation and maintenance ⁴⁰⁵.

5.5 Discussion - The Advance Value Framework

5.5.1 Robustness of the Advance Value Tree

In this paper I developed and proposed a generic model taking the form of a value tree for assessing the value of new medicines in the context of HTA as part of a methodological approach, which moves away from traditional economic evaluation, uses decision theory and adopts multi-criteria evaluation methods. The model consists of a value tree (Advance Value Tree) with three levels (top-level: 5 clusters; middle-level: 10 criteria; bottom-level: 27 sub-criteria or attributes), as shown on Figure 5.2.

Conceptual and theoretical advantages emerge from the methodological and empirical process that was adopted to build the structure of the Advance Value Tree. A five-stage iterative model-building process was followed involving extensive rounds of literature reviews and expert consultation, with the aim to make it as comprehensive as possible whilst maintaining its flexibility and adaptability to serve the needs of decision-makers for different decision problems. A key advantage of the Advance Value Tree is its strong alignment with decision theory principles, adopting a top-down ‘value-focused thinking’ approach while paying attention to the required criteria properties.

For the case of specific decision problems and following the selection of particular treatment options, the value tree would become adapted in a bottom-up approach in order to be able and capture the particular value dimensions of the actual treatment-indication pairs.

This would involve the definition of precise attributes, which should again take place in alignment with decision theory principles and would complete the model-building phase of the process.

Following the model-building phase, model-assessment and model-appraisal phases need to be completed, as part of which the model can be estimated by applying different types of MCDA modelling techniques for the formation of value judgements and elicitation of preferences across the options.

5.5.2 MCDA Methods and Selection of Modelling Techniques to Operationalise the Advance Value Tree

Multiple MCDA methods exist of varying complexity that could be used for the estimation of the model. Specifically, an MCDA modelling technique is required for expressing preferences on the performance of each option against each criterion (scoring), while equating the preference units across all criteria (weighting), and combining preferences of individual criteria together into a combined overall preference (aggregating). Therefore, “modelling” in this context acts as an instrument for enabling decision-makers to understand their own preferences, by helping them to construct their perceptions given a set of assumptions, as part of the overall process of identifying the best decision ².

Among the plurality of existing MCDA approaches, they could be categorized into three main groups of methods governed by different “schools of thought”, notably (a) value measurement methods, including multi-attribute value theory (MAVT) and multi-attribute utility theory (MAUT) methods, (b) outranking methods, and (c) ‘satisficing’ and aspiration level methods as suggested by others ^{1,2,54,92}. These three main groups of methods are essentially referring to different classifications of preference modelling. Other classifications have also been proposed as there is no consensus on a universal categorization ^{195-197,199}.

The methodological process I have proposed pertains to the value measurement methods category ³³², mainly because of the multiple decision contexts that it can be applied to and the simplicity of the value judgements required (in addition to limited restrictions imposed by the axioms employed), features which would probably influence their adoption into the most widely used MCDA methods in health care ⁷². Specifically, I suggest in favour of MAVT methods because of their comprehensiveness and robustness ⁵⁴, as well as their ability to reduce ambiguity and motivational biases. I would therefore suggest the

operationalisation (i.e. estimation) of the Advance Value Tree through such a MAVT approach. This is also consistent with recent work on good practices for MCDA in Health Care Decisions ²⁰⁰. Yet, the value tree and the incorporated criteria clusters could also be operationalised through other MCDA methods that go beyond the value measurement category.

With MAVT methods, value functions for the scoring of the options can be elicited in different ways by mainly using direct or indirect rating techniques ^{2,54}. In contrast to direct rating techniques, indirect techniques aim to uncover decision makers' preferences indirectly through a series of questions that involve differences in the attribute scale and their relation to the value scale. Techniques that explore what is the magnitude of increments in the attribute scale that yield equal units in the value scale are known as indifference techniques, whereas techniques that estimate points on the attribute scale that act as midpoints on the value scale are known as bisection techniques; both of these could be regarded as sub-types of indirect techniques.

Overall, I suggest the use of indirect elicitation techniques, partly because of their unbiased nature given that preferences over the attribute range are first elicited and options then scored indirectly by converting their performance into value scores by producing a transparent valuation relationship across the complete attribute range. For the case of decision contexts requiring repeat decisions, indirect MAVT techniques making use of value functions are recommended in order to ensure the efficient and consistent scoring of alternatives as they become available for evaluation ²⁰⁰, however indirect MAVT techniques could also benefit one-off decisions as they would ensure transparency between the performance of a criterion and the respective value preferences.

In respect to theoretical relevance, choice-based, matching and swing weighting methods are based on MAVT (and MAUT) which need to satisfy a number of axioms for the description of rational choices as described earlier; in contrast, other methods might relax some of these axioms, as in the case of Analytical Hierarchy Process (AHP) that does not assume transitivity of preferences and neither necessarily scales of interval properties but only ordinal. Furthermore, the correspondence of a technique with the analytical requirements (i.e. level of precision) underlying MAVT methods is paramount given that the aim is to estimate a precise value of an option for informing specific decision-making actions ²⁰⁰, conditions which are best met by swing weighting and decompositional techniques; in contrast, in different contexts as in the case of simpler ranking problems, a lower level of correspondence with analytical requirements could be acceptable. Finally, in terms of

cognitive load posed, some techniques might require more data to be considered than others, as for example the simultaneous comparison of multiple criteria required by Discrete Choice Experiments, in contrast to others such as AHP and MACBETH which only consider pairwise comparisons.

Among the variety of MAVT techniques that could be applied to operationalise the Advance Value Tree I would propose the use of “Measuring Attractiveness by a Categorical Based Evaluation Technique” (MACBETH) because of its robustness and simplicity. MACBETH is an approach for the indirect elicitation of value functions through pairwise comparisons of attribute levels and the use of seven word categories to express differences in their value (ranging from “no difference” to “extreme difference”) ^{203,406}. By building a matrix of qualitative judgements of differences in value, it facilitates the move from ordinal preference modelling to cardinal preference modelling, resulting in the development of a quantitative value function model as part of an interactive and constructive approach based on strong theoretical foundations ⁴⁰⁷. As a result, the cognitive load experienced by the decision-makers is at a minimum, offering an easy and comfortable way to express value judgements and elicit preferences, with numerous real world applications illustrating its usefulness as a decision support tool ⁴⁰⁸⁻⁴¹¹. In addition, the use of MACBETH can be explored for deriving and incorporating subjective probabilities through the design and deployment of value - risk matrices which could be used under uncertainty^{412,413}.

As part of weighting, trade-offs between criteria are elicited taking the form of quantitative weights to convert criteria value scores into a common value scale ²⁰⁰. Using direct rating techniques of criteria ‘importance’, as for example, requiring the distribution of 100 points over attributes to reflect their relative importance, is associated with two, potentially serious, problems: first, they could produce flatter importance weight distributions instead of ratio estimation ⁵⁴, which could lead to an underestimation of trade-offs between attributes ⁴¹⁴, and second, they might be insensitive to the attribute ranges used for the performance assessment (i.e. measurement scales) ⁴¹⁵.

Attempting to assign ‘importance’ weights without taking into account the attribute ranges (i.e. measurement scales) are known to be one of the most common mistakes in making value trade-offs ^{3,202}. Instead, the use of indirect swing weighting technique for eliciting relative criteria weights is recommended as common practice through the decision analysis literature ^{54,203,205,416}. This technique involves judgements of relative value between changes (i.e. ‘swings’) from lower performance levels to higher performance levels on each attribute, which is then valued between 0 and 100, the most valuable being anchored at 100

⁵⁴. Alternatively, a Discrete Choice Experiment (DCE), which is based on random utility theory ⁴¹⁷ could be conducted to incorporate a randomness element on responder choices and reflect preference heterogeneity ²⁰⁰. However, the number of criteria would have to be relative small as most DCEs in the field of health care include up to five attributes ⁴¹⁸. Methodological guidance on the sound design and implementation of DCEs taking the form of good research practice is provided elsewhere ⁴¹⁹. Both of these techniques better meet conditions needed in order to treat weights as scaling constants ²⁰⁰, in alignment with decision theory.

Finally, in terms of aggregating, a technique is needed for combining criteria scores and weights together and, more specifically, for selecting a function that allows the combination of the attributes in consistency with responder (e.g. stakeholders) preferences ⁴¹⁴. The application of a simple additive (i.e. linear weighted average) model is the most commonly applied function in health care applications ²⁰⁰, mainly because of its simplicity and comprehensible nature making it easily explained and understood by decision-makers ². However, its use is associated with a number of properties, the most restrictive of which is the existence of preference independence across criteria and attributes, entailing that no two or more criteria or their attributes can independently have a large impact on the overall benefit of the options, in which case a multiplicative or multi-linear model should be used ⁵⁴. Such models are less commonly used though as they are perceived to be more complex to populate ²⁰⁵; additionally, the incorrect aggregation of preferences through such more advanced models can result in considerably greater errors than additive models as evident from empirical evidence of simulation studies ⁵⁷.

Ultimately, results should be examined and sensitivity analysis be conducted, possibly in combination with robustness analysis, in order to validate the model and findings. Deterministic sensitivity analysis can be used to explore the impact of baseline weight (or scores) changes on the rankings of the options and address parameter uncertainty. Robustness analysis can be used as part of an n-way sensitivity analysis to test how simultaneous changes in the criteria scores (or weights) would impact the ranking of the alternative treatments. As a result, differences in viewpoints and any disagreements between stakeholders can be resolved, as for example by testing whether the ranking of the treatments might be sensitive or not to a criterion's relative weight.

5.5.3 *The Advance Value Framework in Perspective*

The application of MCDA modelling techniques for the estimation of the Advance Value Tree through the construction of value judgements and elicitation of preferences based on the MAVT methodological process I have proposed ³³², provides a new value framework based on MCDA principles for the purposes of HTA. The results of the analysis could be used to inform discussions and negotiations on coverage and reimbursement decisions. Overall, this value framework, which I call Advance Value Framework (AVF), adopts an encompassing societal perspective, incorporating views from the wider stakeholder community while assuming that the payer is the ultimate decision maker; however, the choice of perspective could be adapted to different circumstances and decision-making contexts. The AVF can be used for assessing the value of new medicines through the comprehensive set of criteria outlined through the Advance Value Tree; in addition to scientific value judgments relating to therapeutic impact and safety, the value tree allows for the incorporation of social value judgements relating to burden of disease, innovation level and socioeconomic impact, all of which can be captured explicitly.

Finally, the AVF operationalises the Advance Value Tree through the implementation of the combination of (a) an indirect MAVT technique for the elicitation of preferences in the form of value functions (MACBETH), (b) an indirect (qualitative) swing weighting technique, and (c) a simple additive (linear) aggregation approach, altogether producing a decision-making tool easy to be used by all stakeholders, appropriately flexible to meet diverse requirements, and readily adaptable across different settings. I propose this combination of techniques because of their comprehensiveness and methodological robustness, but also their ability of reducing ambiguity and motivational biases. The results of the analysis would be used to inform discussions and negotiations around the coverage and reimbursement decisions.

In reality, the precise choice of scoring, weighting and aggregating techniques will ultimately depend on a number of characteristics of the decision-making problem under consideration, in relation to theoretical relevance, level of precision required in the evaluation of the options and cognitive burden posed to stakeholders and decision-makers ²⁰⁰. Deciding on the optimal combination of modelling techniques represents an important topic that requires further research to better understand the impact of different technique combinations on the above issues and the results of the analysis.

Table 5.2 provides a comparative breakdown of the main features of existing value frameworks and the Advance Value Framework. The development of these value

frameworks is an important step towards a more inclusive Value Based Assessment (VBA) process as part of decision-making contexts in health care. However an ideal framework should be comprehensive enough in terms of the incorporation of value dimensions in order to allow for an adequate capture of value, while on the same time giving the users flexibility for criteria selection based on their specific needs. In any case, individual dimensions involved should possess a number of technical characteristics, if they are to be combined for overall value score rankings to be derived^{204,206,332}.

For example they should be operational so they can be measured and non-overlap so that there is no double counting among them. Importantly, for some frameworks it is not clear how to operationalise performance measurement of the alternative options across the different value dimensions considered (so that options can be assessed) and how to mediate trade-offs among them, and in some cases such efforts seem to lack a theoretical basis, not least because they are derived in an arbitrary manner. Although most of the value frameworks focus mainly on the benefit component of the evaluation process relating to measuring the value of new medicines, issues relating to budget constraints or value for money considerations are crucial to consider in the prioritisation of resource allocation and should be the subject of further research.

5.6 Conclusion

By using a five-stage methodological iterative approach informed by secondary sources and extensive primary research and consultation, I have developed a generic value tree (Advance Value Tree) which is embedded in decision theory. The tree incorporates a number of evaluation criteria that have traditionally been considered in the context of HTA, either explicitly in a systematic manner, or implicitly on an *ad hoc* basis. This work builds on the theoretical foundation of MAVT, based on which the structure of the value tree was derived, influencing the inter-relationship between the different criteria and the extent to which they adhere to a number of critical theoretical properties. I subsequently outlined the assembly of the evaluation criteria in the form of a generic value tree and finally I introduced a number of MCDA methods and proposed a precise combination of techniques for operationalising the value model into a value framework (The Advance Value Framework) for use by decision-makers and stakeholders. In undertaking the above, I focus mainly on the benefit component of the evaluation process relating to measuring the value of new medicines while also accounting for how it could be used in practice given budget constraints in order to

obtain best value for money. Future research could aim to test the value framework in practice, possibly through case studies involving specific interventions, in order to better understand potential advantages and limitations.

Table 5.2: Comparison of the Advance Value Framework with other value frameworks

Framework	ACC/AHA	ASCO	ESMO	ICER	MSKCC	NCCN	MoCA	AVF
Decision context	Clinical practice	Shared decision making	Clinical practice	Coverage/ reimbursement	Pricing	Shared decision making	Pricing and reimbursement	Health Technology Assessment
Key actor	Physicians	Patients - Physicians	Physicians	Payer	Payer-Provider	Patients - Physicians	Payers - Manufacturers	All stakeholders
Value parameters	-Clinical benefit vs. risks -"Value" (cost-effectiveness)	-Clinical benefit (efficacy) -Toxicity (safety) -Cost (efficiency)	-Variability of estimated Hazard Ratio -Observed absolute difference in treatment outcomes:	-Clinical care value -Health system value	-Dollars per life year -Toxicity -Novelty -Cost of development -Rarity -Population burden of disease	-Efficacy of regimen -Safety of regimen -Quality of evidence -Consistency of evidence -Affordability of regimen	-Alternatives available/ unmet need -Relative effectiveness -Response rate -Degree of certainty	-Burden of disease -Therapeutic impact -Safety profile -Innovation level -Socio-economic impact
Conceptual basis	Stakeholder consultation (writing committee)	Stakeholder consultation (ASCO Value in Cancer Care Task Force)	Stakeholder consultation (ESMO Task Force with input from the ESMO faculty and a team of biostatisticians, followed by the ESMO-MCBS	Stakeholder consultation (Input from the ICER Policy Development Group, involving representatives from all major stakeholder groups)	Developed by single clinician/ epidemiologist	Stakeholder consultation (NCCN panel members)	Stakeholder consultation (the MoCA working group that was formed by volunteers from a number of EU countries)	Literature review; Stakeholder consultation (Advance-HTA partners and workshop participants, national experts);

			Task Force, the ESMO Guidelines Committee and a range of invited experts)					Decision theory
Strengths	Quality of evidence (LOE) explicitly ranked; class of recommendation given separately and not averaged together with the level/quality of evidence as a single metric.	Net Health Benefit score and costs illustrated side by side to facilitate the decision making process of patients by making full informed decisions.	Both the variability of the estimated HR and the observed absolute difference in treatment outcomes are explicitly addressed.	Integration of a technology's value-for-money with its' budget impact.	A range of domains incorporated, both relating to the drug and the disease	Easy and simple to comprehend visual output	Easy to comprehend and practical to use because of its simplicity	Multiplicity of explicit value domains; Assignment of quantitative relative weights; Transparent; Engagement of all stakeholders; Grounds on decision theory

Chapter 6 – Paper 4

Application of the MCDA Value Framework: a Simulation Exercise on Metastatic Colorectal Cancer with Multiple Stakeholders in the English Settings¹¹⁹

Summary

Multiple criteria decision analysis (MCDA) has appeared as a possible alternative to address limitations of traditional economic evaluation as part of health technology assessment (HTA), however there is limited empirical evidence from real world applications. The objective of this paper is to test in practice a recently developed MCDA methodological framework through a proof-of-concept case study engaging multiple stakeholders.

A multi-attribute value theory methodological process was used involving consecutive-recursive phases of problem structuring, model building, model assessment and model appraisal. A facilitated decision analysis modelling approach was adopted as part of a decision conference with thirteen participants.

The scope of the National Institute for Health and Care Excellence (NICE) Technology Appraisals setting was used, but, in addition, supplementary evidence was considered for value concerns not explicitly addressed by NICE. Second-line biological treatments for metastatic colorectal cancer (mCRC) patients having received prior chemotherapy were evaluated, including cetuximab monotherapy, panitumumab monotherapy and aflibercept in combination with Folinic acid, (5-) Fluorouracil and Irinotecan (FOLFIRI) chemotherapy. Initially 18 criteria attribute were considered spanning across value domains relating to therapeutic impact, safety profile, innovation level and socioeconomic impact.

Following validation during the decision conference, a total of nine attributes were finally included. The therapeutic impact criteria cluster produced a relative swing weight of 47%, the safety profile cluster 23%, the innovation level cluster 19% and the socioeconomic

¹¹⁹ Chapter 6 has been submitted for peer review with co-authors Prof Gilberto Montibeller, Dr Daniel Hochhauser Prof Panos Kanavos as: Angelis A, Montibeller G, Hochhauser D, Kanavos P. Multiple Criteria Decision Analysis (MCDA) in the Context of Health Technology Assessment: a simulation exercise on metastatic colorectal cancer with multiple stakeholders in the English settings (Under Review at BMC Medical Informatics and Decision Making)

impact cluster 12%. Cetuximab scored the highest overall weighted preference value score of 45.7 out of 100, followed by panitumumab with 42.3, and aflibercept plus FOLFIRI with 14.4. Main limitation was the lack of comparative clinical effects across treatments and challenges included the selection of “lower” and “higher” reference levels on attributes, eliciting preferences across attributes where participants had less experience, and ensuring that all attributes possess the right decision theory properties.

This first application of the Advance Value Framework produced transparent valuations for three mCRC treatments across an explicit set of evaluation criteria proving to aid the decision-making process for participants, however further research is needed to possibly enable its use as part of policy-making.

6.1 Background

The assessment and appraisal of new and expensive medicines by health technology assessment (HTA) bodies, health insurers, and gatekeeper agencies has received considerable attention in recent years, especially in countries with publicly funded health care systems. This is a consequence of negative, and sometime controversial, recommendations on the funding of new medicines due to their high costs. In several cases these medicines relate to treatments for severe diseases with high burden, leading to high patient dissatisfaction and public criticism.

As a result, the methodological aspects for assessing and appraising new medicines have been placed under the microscope. The use of QALYs as part of economic evaluations in HTA, although it is a reasonable measure of health gain it has been argued as inadequate to express the wider patient perspective because it does not reflect other dimensions of social value relating to the burden of the disease, the innovation level of interventions and their wider socioeconomic impact ²¹⁰. These limitations have led sometimes to the *ad hoc* and non-systematic use of additional parameters of value by policy-makers which, in combination with lack of judgements transparency, have given an impression of inconsistency in evidence appraisal and decision-making. Decision controversies however primarily exist because of varying value perspectives, with a disagreement evident among different stakeholders ⁴²⁰. Therefore, for any decision outcomes to be ultimately understood and regarded as “rationally-based”, the application of more comprehensive decision-making processes of an explicit and transparent nature is required.

Developing alternative methodological approaches for the evaluation of new medicines could therefore overcome such limitations, contributing to a more complete

framework for measuring value and making resource allocation decisions. Recently, the use of multiple criteria decision analysis (MCDA) has appeared as a possible alternative to address current limitations of HTA that result from traditional economic evaluation^{90,91,191-194}. Indeed, one of the conclusions of a recent systematic literature review on MCDA approaches applied in health care, including HTA, was that decision-makers are positive about the potential of MCDA to improve decision-making⁷².

However, there are limited studies using empirical evidence from real world MCDA applications with the involvement of stakeholders. In this paper a case study is presented as proof-of-concept, applying a recently developed MCDA methodological framework^{332,421}. A decision conference workshop was organised with the participation of a wide range of stakeholders for evaluating and ranking the value of a set of drugs for the treatment of metastatic colorectal cancer (mCRC) following first line chemotherapy. A facilitated decision analysis modelling approach for expert panels was adopted⁴²². Metastatic colorectal cancer was chosen because of its high severity, the availability of several expensive alternative treatment options, and the fact that it has been the topic of appraisals by several HTA agencies, including a number by the National Institute for Health and Care Excellence (NICE) in England⁴²³⁻⁴²⁸.

The methodological details of the case study are extensively provided in the section below. The overall value rankings of the different drugs are presented in the results section, and the limitations of the study together with the challenges encountered are described in the discussion.

6.2 Methods

6.2.1 Methodological Framework

An MCDA methodological process has been proposed based on Multi-Attribute Value Theory (MAVT)^{54,205} that comprises five distinct phases. These include (a) problem structuring, (b) model building, (c) model assessment, (d) model appraisal, and (e) development of action plans³³². Further detail is provided in *Appendix (6.1)*. The process was operationalised using a decision support system enabling the use of graphics to build a model of values, acting as a facilitation and decision-making tool to inform both the structuring phases (a, b) and the evaluation phases of the process (c, d)⁴²⁹.

6.2.2 Clinical Practice and Scope of the Exercise (Problem Structuring)

This is a simulation exercise focusing on identifying and assessing the overall value of second-line biological treatments for mCRC following prior oxaliplatin-based (first line) chemotherapy, by adopting the respective scope from the latest Technology Appraisal (TA) of each technology that has been appraised by NICE (at the time of study design and data collection, February 2015). In doing so, the same or latest available clinical and economic evidence from the corresponding TAs were used to populate the performance of the alternative options across the respective criteria attributes of our value tree, but in addition supplementary evidence were used for value concerns not addressed by NICE. The scope of TA242 was adopted for the cases of bevacizumab, cetuximab and panitumumab ⁴²⁶, whereas the scope of TA307 was adopted for the case of aflibercept ⁴²⁷. For the case of regorafenib, no sufficient scope details existed in TA334 as the appraisal was terminated early “because no evidence submission was received” from the manufacturer ⁴²⁸, excluding it from the exercise.

The TA242 scoping evaluated bevacizumab in combination with non-oxaliplatin chemotherapy, cetuximab monotherapy or combination with chemotherapy, and panitumumab monotherapy for mCRC after first-line chemotherapy. The populations covered for the case of cetuximab and panitumumab were mCRC patients expressing the wild-type (i.e. non-mutated) form of the v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) gene, because these agents, which target the epidermal growth factor receptor (EGFR), have been shown to be ineffective for treatment of tumours expressing the mutated KRAS gene ⁴³⁰⁻⁴³². KRAS expression does not impact on the case of bevacizumab. Bevacizumab was appraised only when in combination with non-oxaliplatin based chemotherapy, because under UK clinical practice oxaliplatin containing combinations (i.e. FOLFOX) are generally used in the first line. Once cancers demonstrate resistance to FOLFOX patients are then eligible for non-oxaliplatin based chemotherapy regimens such as FOLFIRI, therefore, patients treated with bevacizumab at second-line would normally receive it in combination with non-oxaliplatin based chemotherapies. Hence, the exclusion of oxaliplatin based chemotherapies from the scope of the TA and our exercise.

The scope of TA307 specified the evaluation of aflibercept in combination with FOLFIRI that has progressed following prior oxaliplatin based chemotherapy. Again, the scope of the TA and our exercise considered aflibercept only in combination with non-oxaliplatin chemotherapy for the same reason explained above.

6.2.3 Adaptation of the Advance Value Tree for Metastatic Colorectal Cancer (Model Building)

Overall, a hybrid approach was adopted for the selection of evaluation criteria³³² containing elements both from the “value focused thinking”³ and “alternative focused thinking” approaches²⁰⁴.

A generic value tree (Advance Value Tree) offering an organised overview of the various concerns when evaluating new medicines in an HTA context has been developed under the auspices of the Advance-HTA project¹²⁰ using a combination of literature reviews and expert consultations^{339,421}. The aim was to identify all the necessary criteria for the assessment of value of new medical technologies under a prescriptive approach and it was designed in a top-down ‘value-focused thinking’ manner (criteria selected prior identifying the alternative options)^{3,332}, essentially generating the building blocks of a comprehensive value function. Ultimately, the resulting value tree is decomposed into five value criteria clusters relating to i) the burden of disease the technology addresses (BoD), ii) the technology’s therapeutic impact (THE), iii) the technology’s safety profile (SAF), iv) the overall innovation level (INN) and, v) the wider socioeconomic impact (SOC).

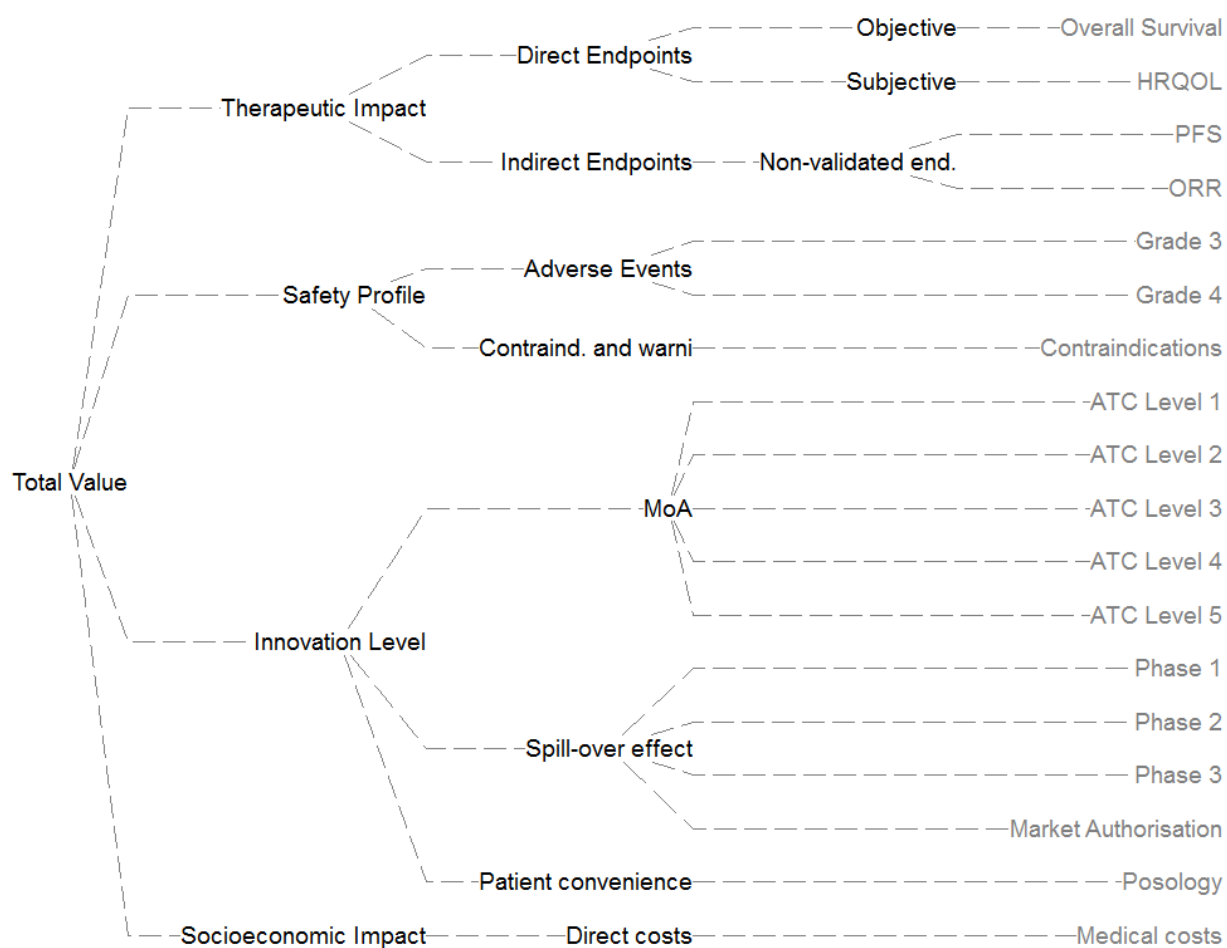
$$Value = f(BoD, THE, SAF, INN, SOC) \quad (6)$$

The generic value tree was later adapted for the context of mCRC in a bottom-up ‘alternative-focused thinking’ approach (criteria emerged following the comparison of the alternative options)^{204,332}. This adaptation resulted in the preliminary version of the mCRC specific value tree (Figure 6.1). Overall, out of the five criteria clusters of the generic value tree, the burden of disease cluster was removed because it was identical across all alternative therapeutic options considered given that all treatments were assessed for the same indication (mCRC). The rest remaining criteria clusters were decomposed into nine sub-criteria clusters with a total of 18 criteria attributes. The list of attributes and their respective definitions are shown in Table 6.1. The selection of the mCRC-specific attributes and consequently the development of the respective value tree took place in alignment with key properties such as preferential independence and non-redundancies in order to ensure their

¹²⁰ Advance-HTA is a research project funded by the European Commission's Seventh Research Framework Programme (FP7). It comprises several complementary streams of research that aim to advance and strengthen the methodological tools and practices relating to the application and implementation of Health Technology Assessment (HTA). It is a partnership of 13 Consortium members led by the London School of Economics - LSE Health.

selection is methodologically correct and theoretically robust according to decision theory principles²⁰⁶.

Figure 6.1: Preliminary value tree for metastatic colorectal cancer (pre-workshop)



*image produced using the Hiview software version 3.2.0.4

Table 6.1: Attributes definition and sources of evidence

<i>Cluster</i>	<i>Attribute</i>	<i>Definition</i>	<i>Evidence source</i>		
			Aflibercept + FOLFIRI	Cetuximab	Panitumumab
THERAPEUTIC IMPACT	Overall survival	The median time from treatment randomisation to death	Van Cutsem et al 2012 ⁴³³	Price et al 2014 ⁴³⁴	Price et al 2014
	HRQoL	Health related uquality of life using EQ-5D score	TA 307 ⁴²⁷	Hoyle et al 2013 ⁴³⁵	Hoyle et al 2013
	Progression free survival	The median survival time on which patients have not experienced disease progression (using RECIST criteria)	Van Cutsem et al 2012	Price et al 2014	Price et al 2014
	Objective response rate	The proportion of patients that experience complete response and partial response (using RECIST criteria)	Van Cutsem et al 2012	Price et al 2014	Price et al 2014
SAFETY PROFILE	Grade 3 AEs	The proportion of patients experiencing a Grade 3 adverse event	Van Cutsem et al 2012	Price et al 2014	Price et al 2014
	Grade 4 AEs	The proportion of patients experiencing a Grade 4 adverse event	Van Cutsem et al 2012	Price et al 2014	Price et al 2014
	Contra-indications	The existence of any type of contraindication accompanying the treatment	EPAR ⁴³⁶ , Prescribing info	EPAR ⁴³⁷ , Prescribing info	EPAR ⁴³⁸ , Prescribing info
	ATC Level 1	The technology's relative market entrance in regards to its ATC Level 1 (Anatomical)	WHO ATC index ⁴³⁹	WHO ATC index	WHO ATC index
	ATC Level 2	The technology's relative market entrance in regards to its ATC Level 2 (Therapeutic)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 3	The technology's relative market entrance in regards to its ATC Level 3 (Pharmacological)	WHO ATC index	WHO ATC index	WHO ATC index

INNOVATION LEVEL	ATC Level 4	The technology's relative market entrance in regards to its ATC Level 4 (Chemical)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 5	The technology's relative market entrance in regards to its ATC Level 5 (Molecular)	WHO ATC index	WHO ATC index	WHO ATC index
	Phase 1	The number of new indications for which the technology is investigated in Phase 1 clinical trials	ClinicalTrials.gov ⁴⁴⁰	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 2	The number of new indications for which the technology is investigated in Phase 2 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 3	The number of new indications for which the technology is investigated in Phase 2 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Marketing authorisation	The number of new indications that the technology has gained an approval for at the stage of marketing authorisation	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Posology	The frequency of doses in a given time period in combination with the duration of the administration	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
SOCIO- ECONOMIC IMPACT	Medical costs impact	The impact of the technology on direct medical costs excluding the purchasing costs of the technology	BNF 69, TA 307, Wade et al 2013 ⁴⁴¹	BNF 69, TA 242 ⁴²⁶ , Hoyle et al 2013	BNF 69, TA 242, Hoyle et al 2013

6.2.4 Evidence Considered and Alternative Treatments Compared (Model Building)

The alternative treatment options compared in the exercise include cetuximab monotherapy (Erbix[®]), panitumumab monotherapy (Vectibix[®]), and aflibercept (Zaltrap[®]) in combination with FOLFIRI chemotherapy. Although there is published evidence for the efficacy of cetuximab in combination with chemotherapy, bevacizumab in combination with non-oxaliplatin-based chemotherapy, or regorafenib monotherapy as treatment options, these treatments were not included in the exercise because there was absence of relevant clinical evidence submitted to NICE as part of their respective TAs^{426,428}.

Overall, evidence sources used to populate the preliminary model included two randomised clinical trials (RCTs)^{433,434}, the respective NICE TAs^{426,427}, any NICE Evidence Review Group (ERG) reports⁴⁴¹ or any related peer review studies coming out^{435,442}, summaries of product characteristics (SPCs) available through EMA's European Public Assessment Reports (EPAR)⁴³⁶⁻⁴³⁸ (or highlights of prescribing information leaflets), Anatomical Therapeutic Chemical (ATC) classification system indexes through the portal of the WHO Collaborating Centre for Drug Statistics Methodology⁴³⁹, and ClinicalTrials.gov listings⁴⁴⁰. The source of evidence used for identifying the performance of options across the criteria attributes is shown in Table 6.1. It should be noted that among the two RCTs used for populating the performance of the treatments across the clinical attributes, one was a head to head trial directly comparing cetuximab versus panitumumab (*ASPECCT* trial)⁴³⁴ and the other one comparing aflibercept in combination with FOLFIRI versus placebo with FOLFIRI (*VELOUR* trial)⁴³³. More details on the clinical evidence considered are shown in *Appendix (6.2)*.

6.2.5 Setting Attribute Ranges and Reference Levels (Model Building)

As part of model building, attribute ranges were selected that were encompassed within minimum (min) and maximum (max) levels. Within the min-max attribute range, "lower" (x_l) and "higher" (x_h) reference levels were defined to act as benchmarks for the preference value scores of 0 and 100 respectively, needed for the construction of criteria value functions and elicitation of relative weights (these are interval scales and thus the importance of setting up clear bounds for each attribute). Incorporation of such intermediate reference levels rather than extreme reference levels at the limits of the scale can protect against inaccuracies emerging from potential non-linearity in value at scale's limits². As a result, value scores could possibly be negative or higher than 100 with $v(x_{lower}) = 0$ and

$v(x_{\text{higher}}) = 100$, essentially conducting a linear transformation which is admissible to an interval scale such as a value scale.

The methodological basis for setting the attribute ranges and the choice of reference levels was the following. For the case of clinical therapeutic attributes, the “higher reference levels” were normally based on BSC figures, coming from the median of the respective arms of the *CO.17* and *AMGEN* trials; otherwise, if no BSC figure was available the placebo comparator arm from the *VELOUR* trial was used. The “lower reference levels” were based on the worst performances plausible, inferred either based on their lowest natural limit (for the case of continuous scale attributes, e.g. 0 months for OS) or based on the lowest evidence-based limit (for the case of non-natural constructed scale attributes, e.g. 0.6 utility for HRQOL as it was the lower utility used for progressive disease by NICE). The maximum levels of the attributes were simply derived by adding a 10% absolute increment to the performance level of the best performing option, essentially offering an error margin to the limits of the scale. This was performed to produce reference levels that corresponded to “worst performance” (plausible) and “satisfactory performance” (proxied by BSC), corresponding to the 0 and 100 anchor levels of the value function scale respectively, with options performing better than the satisfactory level scoring more than 100. By this way three attribute levels were defined in total: i) the “lower reference level” (x_l) (i.e. worst performance plausible), acting on the same time as minimum level (x_*); ii) the “higher reference level” (x_h) (i.e. BSC-based satisfactory performance); and iii) the maximum level (x^*) (i.e. 10% higher than the best performing option), to give $x_* = x_l < x_h < x^*$.

For the purpose of eliciting preferences and producing the matrix of judgements using M-MACBETH, it was aimed to incorporate two intermediate attribute levels lying in-between the three defined attribute levels (giving a total of five different attribute levels) so that the granularity of the scale is increased, essentially to improve the representation of any differences in value across the attribute ranges. In cases where the gaps between the three defined levels were disproportionate large, a third intermediate level was added for a more homogeneous dispersion, giving a total of 6 attribute levels (three defined and three intermediates), whereas in cases of disproportionate small gaps only one intermediate level was added giving a total of 4 attribute levels (three defined and one intermediate). In no cases there were less than 4 and more than 6 attribute levels in total.

Similar but reverse logic was adopted for setting the reference levels of the safety attributes; the “higher reference levels” were based either on the median of the BSC arm from the *AMGEN* trial and the placebo comparator arm of the *VELOUR* trial (BSC data from

the *CO.17* trial were not available for all attributes), or if this was not relevant on the median of the options (e.g. for the case of the existence of contra-indications). The “lower reference levels” were derived either by adding a 10% absolute increment to the worst performing option (e.g. 10% higher incidence for the case of AEs) or by choosing the worst performance plausible for the case of a constructed attribute with a non-continuous scale (i.e. for the existence of contra-indications). The maximum level of the attributes was defined by selecting the best performance plausible (e.g. 0% for incidence of AEs), whereas the minimum level (i.e. worst performance) was equal to the “lower reference level”.

For the innovation attributes, the “higher reference level” was derived by using the median of the options (BSC performance was irrelevant to be used as satisfactory level), whereas the “lower reference level” was based on the worst performance plausible as inferred from the lowest limit of the scales (e.g. 5th entrance at an ATC level, or 0 number of new indications for which the technology is investigated in a given clinical development stage). The maximum level of the attributes was derived by either adding a 10% absolute increment to the performance level of the best performing option, for the case of natural attributes with a continuous scale (e.g. number of new indications for which the technology is investigated in a given clinical development stage), or alternatively by using the best performance plausible for the case of constructed attributes with discrete-level scales (e.g. 1st entrance at an ATC level). The minimum level was equal to the “lower reference level”. For the socioeconomics attribute (impact on direct costs), the “higher reference level” was derived by using the median of the options (BSC performance was irrelevant to be used as satisfactory level), and the “lower reference level” was derived by adding a 10% absolute increment to the worst performing option (i.e. to the one with the biggest impact on costs). The maximum level was defined by selecting the best performance plausible, as inferred from the highest natural limit of the scale (i.e. £0 impact on costs), whereas the minimum level (i.e. worst performance) was equal to the “lower reference level”.

A list of all attributes’ “lower and higher reference levels” together with their basis of selection, as shaped before the workshop is provided in Table 6.2.

6.2.6 Decision Conference (Model Assessment and Appraisal)

The model assessment and model appraisal phases of the exercise took place through a facilitated workshop with key stakeholders and experts, taking the form of a decision conference ⁴⁴³, organised and hosted at the London School of Economics and Political

Science on 30th of April 2015. A decision conference could be defined as “a gathering of key players who wish to resolve important issues facing their organisation, assisted by an impartial facilitator who is expert in decision analysis (DA), using a model of relevant data and judgements created on-the-spot to assist the group in thinking more clearly about the issues” ⁴¹⁶(p.54); see also Franco and Montibeller (2010) ⁴²². Typical stages of decision conference workshops include exploring the issues, structuring and building the model, exploring the model and agreeing on the way forward. In this case, the first two stages were, to a great extent, informed by preparatory work that had been conducted before the workshop, involving extensive literature review.

Background material introducing the scope of the exercise in more detail was sent to the participants one week before the workshop. On the day of the workshop, the model was presented to the participants and was revised cluster by cluster in real time through a facilitated open discussion. It should be highlighted that the aim of the model in this evaluation context is to act as an aid for the group to interact and think about the decision problem, rather than to provide the “correct” answer ^{416,444}. An iterative and interactive model-building process was adopted, where debate was encouraged and differences of opinion actively sought. Generally, overall agreement was reached in regards to criteria inclusion and exclusion; in the few instances where this was unattainable, criteria were left in the model for their impact to be tested as part of the sensitivity analysis stage, where distinctive viewpoints were finally resolved.

Table 6.2: Pre-workshop attribute reference levels and basis of selection

Cluster	Attribute	Metric	Lower level	Basis	Higher level	Basis
THERAPEUTIC IMPACT	Overall survival	months	0	Minimum limit of the scale	6.2	BSC
	HRQoL	utility score (EQ-5D)	0.6	Lower score used for progressive state in TA307 ⁴²⁷	0.75	BSC
	Progression free survival	months	0	Minimum limit of the scale	1.9	BSC
	Objective response rate	% of patients	0	Minimum limit of the scale	11	FOLFIRI + Placebo (VELOUR trial) ⁴³³
SAFETY PROFILE	Grade 3 AEs	% of patients	68	10% higher than the worst performing option	32	Median of BSC (AMGEN trial) ⁴⁴⁵ and FOLFIRI + Placebo (VELOUR trial)
	Grade 4 AEs	% of patients	24	10% higher than the worst performing option	10	Median of BSC (AMGEN trial) and FOLFIRI + Placebo (VELOUR trial)
	Contra-indications	types of contra-indications	Lower expected benefit and higher expected risk	Minimum limit of the scale	Lower expected benefit	Median of options
INNOVATION LEVEL	ATC Level 1	relative market entrance	5th	Minimum limit of the scale	4th	Median of options

	ATC Level 2	relative market entrance	5th	Minimum limit of the scale	4th	Median of options
	ATC Level 3	relative market entrance	5th	Minimum limit of the scale	3rd	Median of options
	ATC Level 4	relative market entrance	5th	Minimum limit of the scale	1st	Median of options
	ATC Level 5	relative market entrance	5th	Minimum limit of the scale	1st	Median of options
	Phase 1	number of new indications	0	Minimum limit of the scale	17	Median of options
	Phase 2	number of new indications	0	Minimum limit of the scale	55	Median of options
	Phase 3	number of new indications	0	Minimum limit of the scale	18	Median of options
	Marketing authorisation	number of new indications	0	Minimum limit of the scale	2	Median of options
SOCIO-ECONOMIC IMPACT	Posology	duration of administration & frequency of doses	Many hours, every 2 weeks	Minimum limit of the scale (worst performing option)	Up to an hour, every 2 weeks	Maximum limit of the scale (best performing option)
	Medical costs impact	GBP (£)	7,086	10% higher than the worst performing option	4,589	Median of options

The composition of the group's expertise and the numbers of the different stakeholders were decided based on the structure of the past NICE committees responsible for the appraisals of the alternative treatments ^{423,425-427}. It was aimed to involve a small group between 7 and 15 participants; these group sizes have been shown to be adequate because they tend to preserve individuality while also allowing efficient group processes, as they are small enough to be able to work towards agreement, but large enough to represent all major perspectives ⁹⁴. In total, 13 participants were involved, their areas of expertise and type of affiliation are shown in Table 6.3.

Participants were contacted through an email invitation outlining the exercise and the purpose of the project, and background material introducing the scope of the exercise in more detail was sent one week before the workshop. Travel expenses for the participants were retrospectively covered.

In terms of the decision-aiding methodology used, impartial facilitation was adopted with the aim of enhancing content and process interaction, while refraining from contributing to the content of the group's discussions, essentially guiding the group in how to think about the issues but not what to think ^{416,446}. In terms of facilities, the room of the workshop had a U-shaped meeting table for all the participants to have direct eye contact, with an overhead projector screen surrounded by whiteboards. The M-MACBETH software was operated using a laptop, the screen of which was connected to the projector.

The workshop lasted the whole day, from 9.00 am to 18.00 pm with one 45-minutes lunch break, and two 15-minutes coffee breaks. The day started with a brief introductory presentation and then moved on with an overview of the MCDA methodology adopted and the description of the value tree. The value tree was then presented and analysed cluster by cluster.

At the beginning of each cluster the value tree was validated; the various criteria were explained, followed by a group discussion relating to their relevance and completeness. As a result of this iterative process, some of the criteria were excluded because they were perceived as irrelevant or non-fundamental, but no criteria were deemed to be missing. Then, value functions were elicited for the different criteria and criteria weights were elicited within the clusters. Finally, relative weights were assigned across clusters, which enabled calculating the overall WPV scores of the options.

Table 6.3: List of decision conference participants

Participant	Expertise	Affiliation
1	Medical oncologist - CRC expert	NHS Trust - Teaching hospital
2	Medical oncologist - CRC expert	NHS Trust
3	Consultant - community paediatrician	NHS Trust - HTA agency
4	Public health expert	Academia
5	Pharmacist	Independent
6	Health economist	Academia
7	HTA expert	Academia
8	Health economist	Academia
9	HTA expert	Academia
10	Medical statistics	Academia
11	Patient	Independent
12	Patient carer	Independent
13	Patient advocate	Charity

6.2.7 MCDA Technique (Model Assessment and Appraisal)

The model was operationalised through a value framework consisting of a value measurement method adopting a typical simple additive aggregation approach, where the overall value $V(.)$ of an option a is given by Equation 4⁴²¹. Overall, the additive value model is assumed as a working hypothesis and therefore the model construction is undertaken to respect its underlying properties (see section 1.1.4), requiring positive weights summing one and the use of explicit reference values of 0 and 100, see Table 6.2 and 6.4.

Value preferences can be elicited using different protocols. A method based on pairwise qualitative comparisons is MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique), an approach using qualitative judgements about the difference of attractiveness between different pairs of attribute levels^{406,447}. Under MACBETH, which based on strong theoretical foundations⁴⁰⁷, semantic judgements are converted into a cardinal scale therefore providing a simple, interactive and constructive approach, and its usefulness as a decision support tool has been shown through numerous applications for real world problems⁴⁰⁸⁻⁴¹¹.

M-MACBETH, a decision support system based on the MACBETH approach ⁴⁴⁸, was used to elicit value preferences of the workshop participants and more precisely to build the value tree, elicit the value functions for the different attributes, assign relative attribute weights through a qualitative swing weighting approach, aggregate the preference value scores and weights using an additive aggregation (i.e. simple additive model) to derive overall weighted preference value (WPV) scores, and conduct sensitivity analysis ⁴⁴⁹. Besides a consistency check between the qualitative judgements expressed that is automatically provided by the software, a second consistency check was performed manually to ensure that an interval scale is obtained, i.e. validate the cardinality of the scale, by comparing the sizes of the intervals between the suggested scores and inviting participants to adjust them if necessary ⁴⁵⁰, an essential requirement for aggregation using simple additive value models. More technical details on MACBETH are provided in *Appendix (6.3)*.

6.2.8 Costs Calculation (Model Assessment and Appraisal)

Drug costs were calculated according to prices (excl. VAT), pack sizes and dosage strengths as found on the British National Formulary (BNF 69), and the recommended dosage and treatment duration as reported on the respective NICE technology appraisals^{426,427}. Vial wastage was assumed in all calculations. Drug administration costs for cetuximab and panitumumab were kept consistent with Hoyle et al. ⁴⁴² and administration costs for aflibercept plus FOLFIRI consistent with the respective ERG Report ⁴⁴¹.

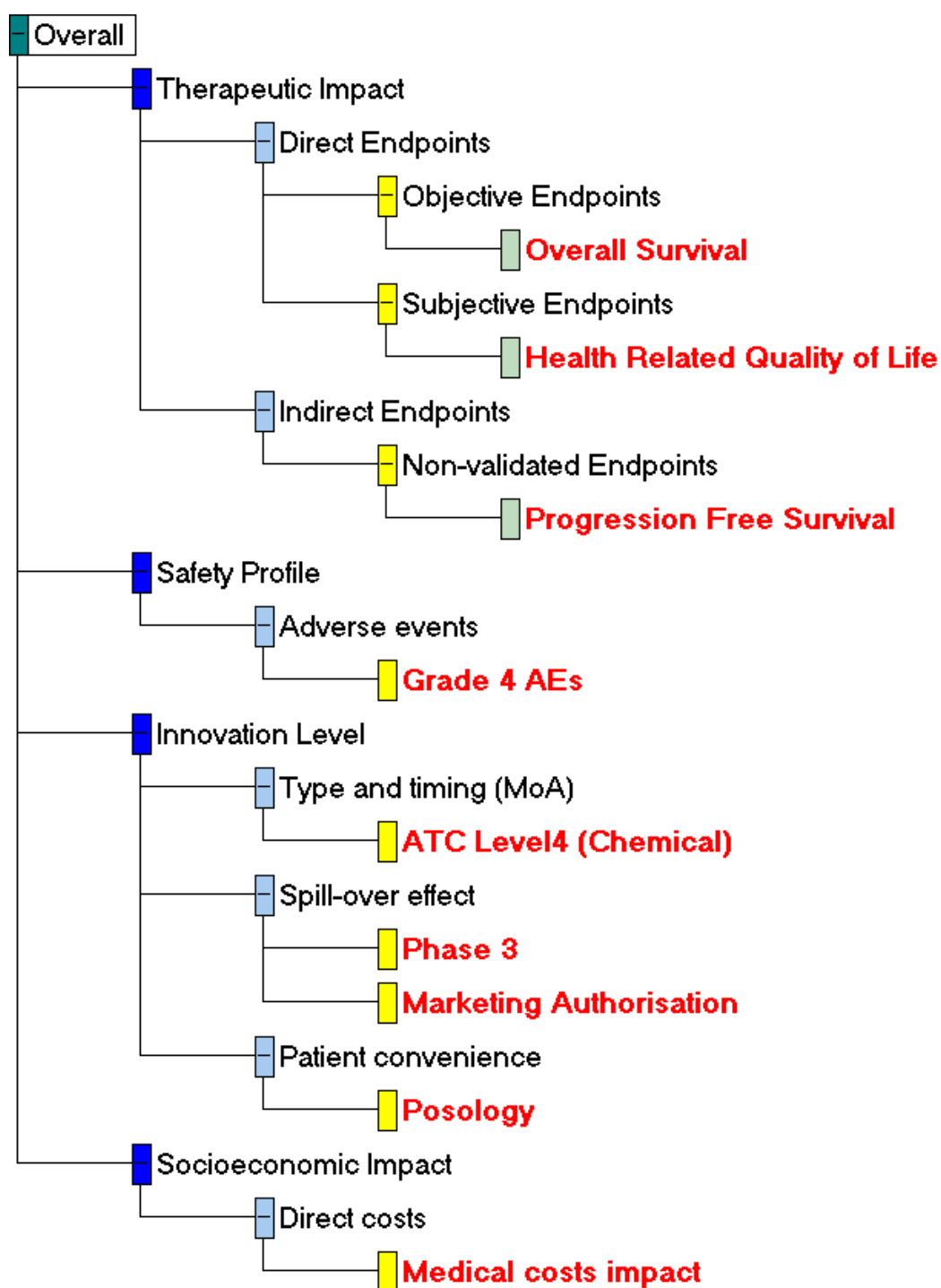
6.3 Results

6.3.1 Criteria Validation and Amended Value Tree for Metastatic Colorectal Cancer

The final version of the value tree, as emerged following the open discussion with the participants in the workshop is shown in Figure 6.2. In total, 9 out of the 18 attributes were removed from the value tree because they were judged from the participants to be non-fundamental for the scope of the exercise, resulting in a value tree with half of its' original size. Importantly, no criteria were deemed to be missing. In the therapeutic impact cluster, the Objective Response Rate (ORR) attribute was removed. In the safety profile cluster, the contra-indications attribute was removed and the Grade 3 Adverse Events (AEs) and Grade 4 AEs attributes were proposed to be aggregated into a single attribute; however this aggregation required a significant modelling iteration, and due to time constraints it was decided to exclude the Grade 3 AEs attribute and only include Grade 4 AEs for the purpose

of the simulation exercise. In terms of the innovation cluster, participants had mixed views. Consensus (i.e. full agreement) was reached for the ATC L5, Phase 1 and Phase 2 attributes to be removed, and for the ATC L4 to be included as a binary variable (i.e. first entrance in the chemical class vs. second or subsequent entrance in the chemical class); however strong disagreement existed on whether to include Phase 3 and Market Authorisation attributes, with half of the participants in favour and half of the participants opposed to their inclusion. As a result, both of the attributes were left in the model and their impact was then tested at the end of the workshop as part of the sensitivity analysis stage.

Figure 6.2: Final Value Tree for metastatic colorectal cancer (post-workshop)



* image produced using the M-MACBETH (beta) software version 3.0.0

6.3.2 Validation of Attribute Ranges and Reference Levels

Another important amendment in the model included a change in the definitions of the “lower” and “higher” reference levels, which define the 0 and 100 scores in the value functions and anchor the swing weights. For the case of clinical attributes a majority agreement was reached by the group’s participants that the “lower reference level” should actually correspond to the “satisfactory performance” (proxied by Best Supportive Care, BSC) rather than the “worst performance” plausible. As a result, the “lower reference level” was switched to the previously defined “higher reference level” (i.e. satisfactory performance), and the “higher reference level” was set equal to the maximum level. The newly-defined attribute levels were therefore: i) the “lower reference level” (i.e. BSC-based satisfactory performance); the “higher reference level” (equal to the maximum, best performance plausible, level), and the minimum level (i.e. worst performance plausible). In doing so, the amended “lower and higher reference levels” were now corresponding to “satisfactory performance” (proxied by BSC) and “best performance” respectively, with options performing worse than the “satisfactory” level getting a negative score, and all options obtaining less than 100 score.

As a consequence, a similar change was introduced for the case of - the now single - safety profile attribute (Grade 4 AEs), with the “lower reference level” being defined based on “satisfactory performance”, the “higher reference level” based on “best performance” (i.e. minimum limit of the scale) and the minimum level remaining the same.

For the case of innovation attributes, the “higher reference levels” were set equal to the “best performance” levels, with the “lower reference levels” remaining the same (equal to the worst performance). Similarly, for the case of the socioeconomic impact attribute (impact on direct medical costs) the “higher reference level” was also set equal to the “best performance” level, with the “lower reference level” remaining the same (equal to the worst performance).

The arising changes in the attribute reference level definitions, before and after the workshop for each of the criteria clusters are shown in Table 6.4, with the final list of attributes’ “lower” and “higher” reference levels, together with their basis of selection provided in Table 6.5.

Table 6.4: Changes in the definitions of the attribute reference levels, pre- and post-workshop

Attribute level	Preliminary (pre-workshop)	Final (post-workshop)
Overall		
<i>Maximum</i>	best performance ²	best performance ²
<i>"Upper reference"</i>	satisfactory performance ²	best performance ²
<i>"Lower reference"</i>	worst performance	satisfactory performance ²
<i>Minimum</i>	worst performance	worst performance
Therapeutic Cluster		
<i>Maximum</i>	10% > best performing option	10% > best performing option
<i>"Upper reference"</i>	BSC performance	10% > best performing option
<i>"Lower reference"</i>	worst performance	BSC performance
<i>Minimum</i>	worst performance	worst performance
Safety Cluster		
<i>Maximum</i>	best performance/ limit of scale	best performance/ limit of scale
<i>"Upper reference"</i>	BSC performance (or median performance of options)	best performance/ limit of scale
<i>"Lower reference"</i>	10% > worst performing option (or worst performance)	BSC performance (or median performance of options)
<i>Minimum</i>	10% > worst performing option (or worst performance)	10% > worst performing option (or worst performance)
Innovation Cluster		
<i>Maximum</i>	10% > best performing option or best performance/ limit of scale	10% > best performing option or best performance/ limit of scale
<i>"Upper reference"</i>	median performance of options	10% > best performing option or best performance/ limit of scale
<i>"Lower reference"</i>	worst performance/ limit of scale	worst performance/ limit of scale
<i>Minimum</i>	worst performance/ limit of scale	worst performance/ limit of scale
Socioeconomic Cluster		
<i>Maximum</i>	best performance	best performance
<i>"Upper reference"</i>	median performance of options	best performance
<i>"Lower reference"</i>	10% > worst performing option	10% > worst performing option
<i>Minimum</i>	30% > worst performing option	30% > worst performing option

Table 6.5: Post-workshop attribute reference levels and basis of selection

Cluster	Attribute	Metric	Lower level	Basis	Higher level	Basis
THERAPEUTIC IMPACT	Overall survival	months	6.2	BSC	14.9	10% higher than the best performing option
	HRQoL	utility score (EQ-5D)	0.75	BSC	0.9	10% higher than the best performing option/ general population
	Progression free survival	months	1.9	BSC	7.6	10% higher than the best performing option
SAFETY PROFILE	Grade 4 AEs	% of patients	10	Median of BSC arm from AMGEN trial ⁴⁴⁵ and placebo + FOLFIRI arm from VELOUR trial ⁴³³	0	Maximum limit of the scale
INNOVATION LEVEL	ATC Level 4	relative market entrance	≥2nd	Minimum limit of the scale, binary variable	1st	Maximum limit of the scale, binary variable
	Phase 3	number of new indications	0	Minimum limit of the scale	21	10% higher than the best performing option

	Marketing authorisation	number of new indications	0	Minimum limit of the scale	3	10% higher than the best performing option
	Posology	duration of administration & frequency of doses	Many hours, every two weeks	Minimum limit of the scale (worst performing option)	Up to an hour, every two weeks	Maximum limit of the scale (best performing option)
SOCIO-ECONOMIC IMPACT	Medical costs impact	GBP (£)	7,086	10% higher than the worst performing option	0	BSC

6.3.3 Options Performances, Criteria Weights and Overall Preference Value Rankings

Two examples of value judgements matrices and their conversion into a linear and non-linear value function for the case of the Overall Survival (OS) and Health Related Quality of Life (HRQoL) attributes respectively are shown in *Appendix* (6.4), together with the value functions for the rest attributes.

The performance of the options across the different attributes together with the “lower” and “higher” reference levels is shown in Table 6.6. The different columns correspond to the performance of the different options (including the two reference levels), across the respective attributes as shown across the rows. The table of overall WPV scores for all options across the different attributes, together with the respective attribute baseline weights is shown in Table 6.7; similarly to Table 6.6, the different columns correspond to the preference value scores of the different options (including the two reference levels), across the respective attributes as shown across the rows. Cetuximab scored the highest overall WPV score of 45.7, followed by panitumumab with an overall WPV score of 42.3. Afibercept in combination with FOLFIRI was ranked last with an overall WPV score of 14.4, partially due to its performance on Grade 4 AEs (21%) which lied below the lower reference level of the value scale (10%), producing an absolute preference value score of -117.9 and a weighted preference value score of -27.4. A stacked bar plot of the overall WPV scores of the alternative treatments across the attributes is shown in Figure 6.3.

The relative weights assigned to the different attributes are shown in Figure 6.4. The criteria are ranked based on their relative magnitude, ranging from the relatively more important criteria to relatively less important criteria (from left to right across the x-axis), taking into account the “lower” – “higher” ranges of the attributes. The OS and Grade 4 AEs attributes together assigned a relative weight totaling the relative weights of all other attributes together, i.e. 50%. Out of 100, the therapeutic impact cluster attributes totaled overall a relative weight of 47, the safety profile cluster (single attribute only) a relative weight of 23, innovation level cluster attributes totaled overall a relative weight of 19, and the socioeconomic impact cluster (single attribute only) a relative weight of 12.

Table 6.6: Options performance across the criteria attributes

Attribute name	Attribute metric	Lower level	Aflibercept + FOLFIRI	Cetuximab	Panitumumab	Higher level
Overall Survival	months	6.2	13.5	10	10.4	14.9
HRQoL	utility (EQ-5D)	0.75	0.78	0.78	0.78	0.9
Progression Free Survival	months	1.9	6.9	4.1	4.4	7.6
Grade 4 AEs	% of patients	10	21	5	7	0
ATC L4	relative market entrance	2nd	1st	1st	2nd	1st
Phase 3	# of new indications	0	18	19	7	21
Marketing Authorisation	# of new indications	0	3	1	0	3
Posology	duration & frequency	hrs, every 2 weeks	hrs, every 2 weeks	1 hr, every week	≤hr, every 2 weeks	≤hr, every 2 weeks
Medical costs impact	GBP (£)	7,086	6,738	4,589	1,940	0

Table 6.7: Overall weighted preference value (WPV) scores, partial preference value scores, relative weights, costs and cost per unit of value

	Lower level	Aflibercept + FOLFIRI	Cetuximab	Panitumumab	Higher Level	Relative Weights
Overall WPV score	0.0	14.4	45.7	42.3	100.0	100
Overall survival	0.0	83.9	44.4	48.9	100.0	29
Health Related Quality of Life	0.0	15.0	15.0	15.0	100.0	13
Progression free survival	0.0	90.3	51.4	55.6	100.0	5
Grade 4 AEs	0.0	-117.9	50.0	30.0	100.0	23
ATC L4	0.0	100.0	100.0	0.0	100.0	6
Phase 3	0.0	50.0	66.7	19.4	100.0	2
Marketing Authorisation	0.0	100.0	30.0	0.0	100.0	3
Posology	0.0	0.0	37.5	100.0	100.0	7
Medical costs impact	0.0	7.0	50.0	78.9	100.0	12
<i>Costs (£)</i>		29,400	18,000	27,000		
<i>Cost per unit of value</i>		2,046	394	638		

Figure 6.3: Stacked bar plot of treatments’ overall weighted preference value (WPV) scores across all attributes

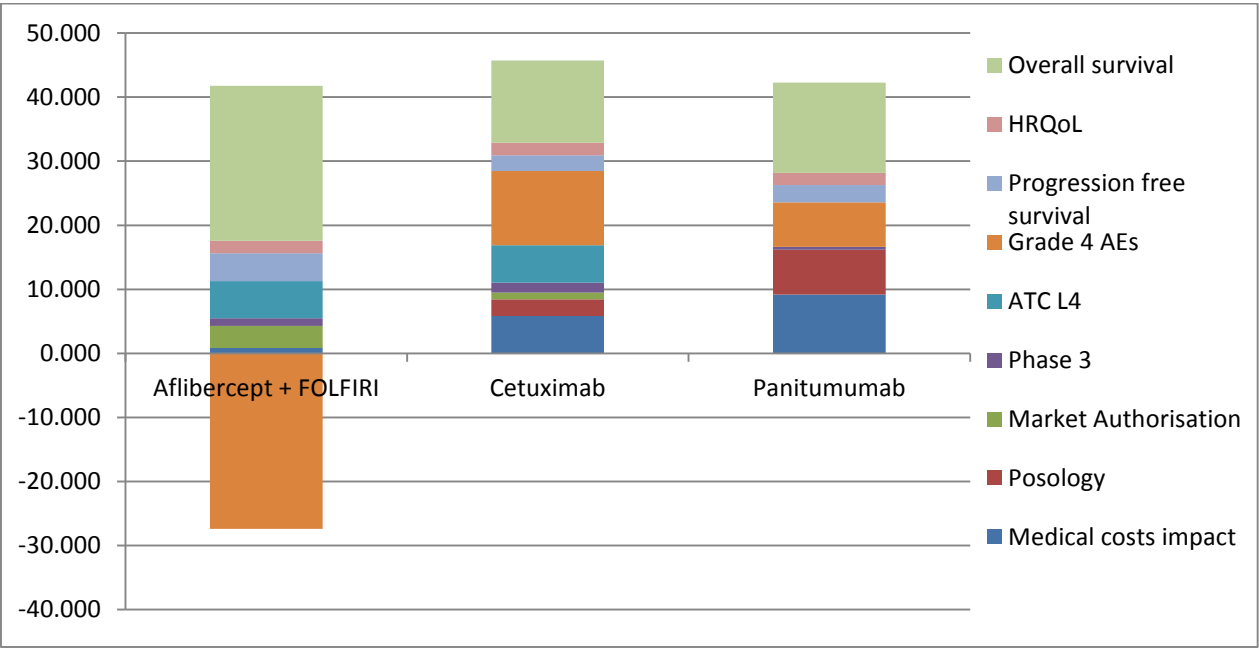
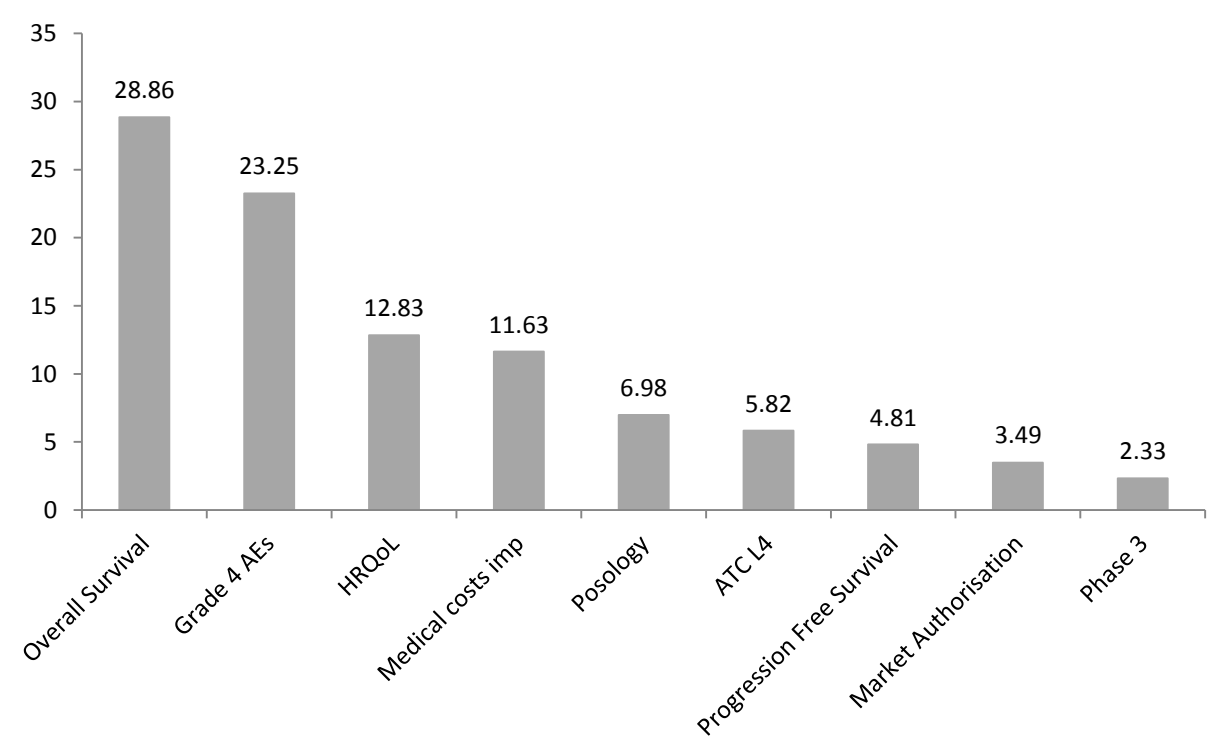


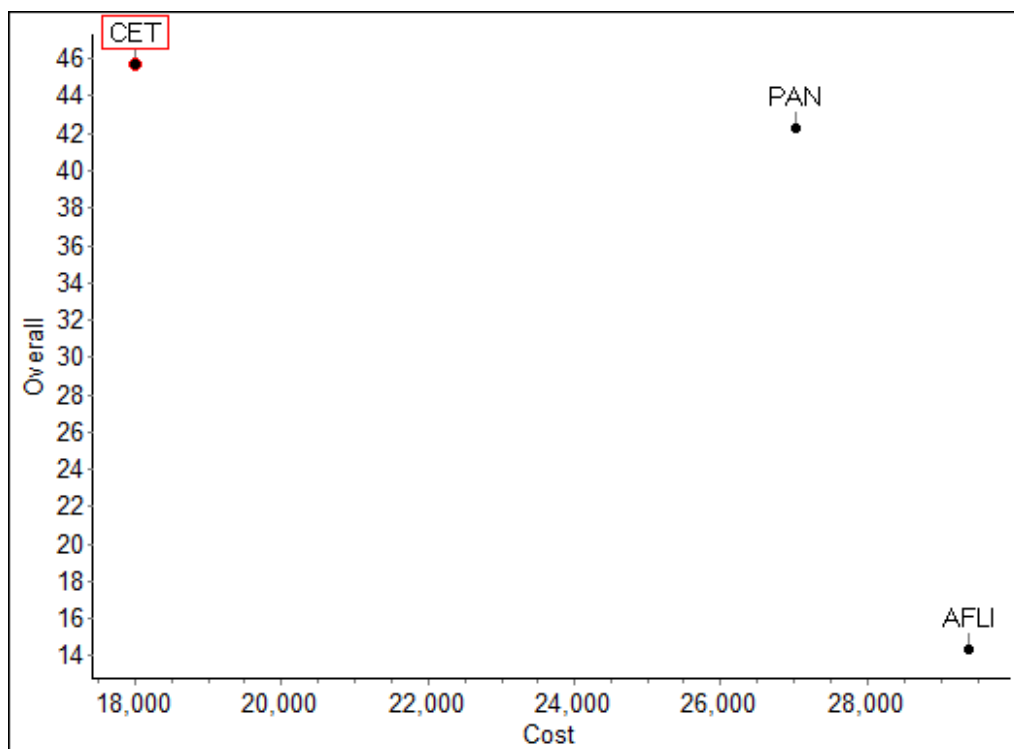
Figure 6.4: Criteria weights histogram



6.3.4 Value for Money Analysis

By incorporating the total purchasing costs of the different drugs (including their administration costs), their overall WPV scores versus total costs plot is produced (Figure 6.5). By using rounded up total cost figures of £18,000 for cetuximab (£12,824 drug cost and £5,191 administration cost), £27,000 for panitumumab (£23,643 drug cost and £3,374 administration cost), and £29,400 for aflibercept in combination with FOLFIRI (£17,750 drug cost and £11,630 administration cost), and dividing them with overall WPV scores, their costs per MCDA value unit were calculated to be £394, £638, and £2,046 respectively (Table 6.7). Therefore, in terms of value-for-money, aflibercept in combination with FOLFIRI is shown to be dominated by panitumumab, both of which are shown to be dominated by cetuximab which is associated with the highest overall WPV score and the lowest cost.

Figure 6.5: Cost benefit plot of overall weighted preference value (WPV) scores versus costs



* image produced using the M-MACBETH (beta) software version 3.0.0

6.3.5 Sensitivity and Robustness Analysis

Deterministic sensitivity analysis was conducted to address parameter uncertainty by exploring the impact of baseline weight changes on the ranking of the options (figures shown in *Appendix 6.5*). In order for panitumumab to become better ranked than cetuximab any of the following changes in baseline weights would be needed: OS from 28.9 to 59.8, PFS from 4.8 to 47.7, Grade 4 AEs from 23.3 to 7.5, ATC L4 from 5.8 to 2.5, Posology from 7 to 11.8, or Medical costs impact from 11.6 to 21.0.

Similarly, for aflibercept plus FOLFIRI to become better ranked than cetuximab, OS weight from 28.9 to 60.3, PFS from 4.8 to 47.3, Grade 4 AEs from 23.3 to 5.6, or Market authorisation from 3.5 to 33.3. Finally, for aflibercept plus FOLFIRI to become better ranked than panitumumab, OS weight from 28.9 to 60.4, PFS from 4.8 to 47.2, Grade 4 AEs from 23.3 to 5.4, ATC L4 from 5.8 to 26.4, Phase 3 from 2.3 to 49.0, or Market authorisation from 3.5 to 24.6.

Therefore, conclusions were fairly robust as treatment rankings were not influenced by changes of at least 50% or less on any of the baseline normalised weights, the most sensitive attributes being Posology and Medical costs impact attributes on the cetuximab versus panitumumab comparison (requiring a 69% and 81% change respectively for panitumumab to become better ranked), with changes of at least up to 100% on the remaining baseline normalised weights exerting no impact on the results.

The robustness of the results was also tested by conducting 8-way sensitivity analysis in the reference levels of the attributes using the respective function of the M-MACBETH software (“Robustness analysis”), which showed that a simultaneous change of up to 5% across all of the attribute reference levels would not impact the ranking of the alternative treatments (figure in *Appendix 6.5*).

However other types of uncertainty might exist, such as stochastic uncertainty, structural uncertainty and heterogeneity, which could be addressed through more advanced statistical approaches, including probabilistic sensitivity analyses, Bayesian frameworks, fuzzy set theory or grey theory²⁰⁸. For example, uncertainty associated with the performance of the options due to sampling variation of clinical studies, or with the criteria weights due to inability to derive or agree on weights, might make the application of point estimates inappropriate in which case stochastic multi-criteria acceptability analysis (SMAA) might be preferred⁴⁵¹.

6.4 Discussion

This was a simulation exercise following an MCDA methodological process adapted for the context of HTA ³³² and testing a newly developed MCDA value framework in practice through a decision conference ⁴²¹. The methodological process adopted is generally in alignment with recent good practice guidelines on the use of MCDA for health care decision-making, in respect to design, implementation and review of the analysis ²⁰⁰.

Overall, a set of different treatment options for the indication of mCRC at second-line were assessed and ranked based on their overall WPV scores. These scores acted as value metrics or value indices, comprised from the performance of the alternative treatment options against an explicit set of criteria while adjusting for the relative importance of these criteria, as reflected by the preferences of the group. Finally, incorporation of drug costs (purchasing and administration costs) enabled the production of “cost per unit of value” ratio estimates while revealing the dominance of one treatment.

Assuming that the participants of the workshop acted as a group responsible for choosing or recommending the funding of one of the alternative options, cetuximab would be the rational choice of the group assuming that the respective budget needed is available. Incorporation of budget impact considerations in the cost-value ratio estimates at system level could take place by taking into account the number of patients that will receive the treatment, however benefits and costs should be estimated in comparable units so that the results are not biased towards the cheapest alternative, as for example on a per-patient basis ²⁰⁰.

Participants’ feedback about the case study highlighted the importance of having a fully transparent evaluation model that can be used to construct and analyse value preferences, including value trade-offs, and pointed towards the existence of promising prospects with the use of the adopted methodology in the future.

6.4.1 Strengths and Opportunities

Among the biggest benefits of the methodology adopted is the explicit incorporation of multiple benefit dimensions, some of which are possibly hard-to-measure, but proved important nevertheless.

Another central strength of the framework is the development of the evaluation model with a group of relevant stakeholders (health care professionals, methodology experts, patients), which proved to be essential for creating a shared understanding of what

constitutes value in this decision context. This was evident across all its phases, ranging from model-building to model appraisal, playing a profound role across all the stages. Starting with criteria selection, by sharing participants' views and opinions among the group while seeking a consensus (i.e. full agreement if not majority agreement) approach, the original version of the value tree and its criteria were validated, amending its contents and leading to the exclusion of some attributes that seemed non-fundamental or irrelevant. For example *Objective Response Rate (ORR)*, the sum (i.e. proportion) of patients that experience complete response and partial response (using the RECIST criteria), which was originally included in the value tree was decided to be removed because of irrelevance. Initially, the clinical view was raised that stable cancer (i.e. non-responding) might be just as good of an outcome as tumour shrinking. Although the argument was expressed that in theory *ORR* could help into controlling symptoms better, there is no firm evidence for this in the literature. The *HRQoL* and *Grade 4 AEs* indices are instruments designed for the assessment of symptoms and *ORR* gives no additional utility. Thus, its inclusion in this regard would even entail double counting effects. In turn, it was suggested that *ORR* is primarily designed for measuring patient response and treatment efficacy under the settings of new drug development and not a major guide to clinical practice in the setting of advanced disease. With regard to clinical practice, the use of *PFS* as a metric could be perceived more complete and reflective.

During model assessment and the elicitation of preferences through value functions, a structured discussion enabled the representation of all the different perspectives for the purpose of valuation. Although occasionally some of the participants might at first have had opposing views and beliefs in regards to their preference judgments, in most of the cases these conflicts were terminated or defused following extensive discussions. An example would be the elicitation of the *Overall Survival (OS)* value function which started with contrasting perspectives on how to assess additional months of life, but following far-reaching dialogues around the added value of different life increments, an agreement was established that each additional month of life was associated with an equal magnitude of value, as revealed through a linear value function.

The systematic assessment of all types of evidence together enabled the identification of strengths and weaknesses for each treatment, which in turn could be used to influence their use under clinical practice, or even support their redesign and improvement as part of clinical development process. For example, although aflibercept in combination with FOLFIRI (afli+FOL) was associated with the highest score in *OS*, overall it ranked last

because of a highly negative score in *Grade 4 AEs*. Given that AEs might be correlated with the dose of the treatment, and considering the preferences of the group, it could thus be suggested that a lower dose of afli+FOL could potentially produce a lower incidence of *Grade 4 AEs* at possibly acceptable reductions in *OS* as a trade-off (given the relatively higher score of the option on the latter attribute), which could improve its overall WPV score. Importantly, the clinical view was raised that in practice the dosage of afli+FOL administered to patients might be higher than what prescribing guidelines recommend, which would support the previous argument. However in reality a lower dose might give rise to much lower efficacy/effectiveness (not being linearly correlated), and this is strictly a hypothetical assumption aiming to give an insight on how the results of such an analysis could be taken into consideration.

In turn, the negative score of afli+FOL on *Grade 4 AEs* was directly influenced by the definitions of the “higher” and “lower” reference levels, as outlined on Table 6.5. One could argue that setting the “higher reference level” on the basis of the minimum limit of the scale (i.e. zero) might be unrealistically idyllic. As part of a secondary post-workshop analysis the “higher” reference level of the *Grade 4 AEs* attribute was changed from 0% (best possible performance, minimum limit of the scale), to 2% (BSC performance) in order to test the impact on the score of the treatments, which resulted to an even higher negative score for afli+FOL (this is because the other options were now scoring higher, as their performance was located closer to the “higher” reference level). Assuming that such an analysis and discussion were conducted at an earlier stage of the drug’s life cycle, for example during Phase 2 clinical trials, these insights could instead influence future aspects of drug development by inducing changes into the formulation process and the envisaged posology characteristics of the drug.

Another benefit of the evaluation was a clear separation between the performance of treatments and their valuation, based respectively on the availability of evidence across the attributes and the establishment of marginal value within criterion and value trade-offs across criteria, with the latter one being amenable to sensitivity analysis. The explicit modelling of preferences and values represented how much the group valued the marginal performances for each attribute, as well as their priorities for the different criteria, represented by the weights. This separation allowed assessing the robustness of results for variations of preferences. For example, sensitivity analysis at the end of the workshop in respect to the baseline weights of the innovation attributes, for which some of the participants did not totally agree with their elicited relative importance, assured the participants that the

ranking of the treatments was not sensitive to minor variations along their range (last column of Table 6.7).

6.4.2 Limitations and Challenges

Among the limitations of the case study was the use of un-synthesised evidence to inform the clinical attributes given the lack of relative treatment effects, therefore limiting the extent to which the results can inform real policy-making. An important challenge would relate to the technical difficulties in ensuring that all attributes possess the required theoretical properties, and more specifically the fact that potential preference-dependence was observed between the OS and HRQOL attributes which was addressed by deriving OS value functions that were effectively conditional on the range of the HRQOL attribute. However, it should be noted that the use of conditional value functions might be incompatible with the use of an additive model and therefore should only be used with caution, as for example within descriptors of performance (i.e. attributes). In order to be able to use an additive model under the existence of preference dependence, the effect between the two attributes would have to equal zero, or, in other words the product of their value scores would have to be zero; for a more extensive discussion around the use of linear additive model together with conditional value functions due to preference dependence see section 8.3.2. Another limitation would be that the HRQOL of the progressive disease was not assessed because none of the treatments assumed to have any effects on it, something which might not hold true in other conditions. Among the main challenges included the evaluation of attributes in which participants had less experience which was addressed from expert opinion. These limitations and challenges are extensively discussed in the Limitations and Challenges section in the last chapter of the thesis (section 8.3.2).

6.5 Conclusion

The challenge to assess novel treatments and therapeutic combinations in a setting of significant budgetary pressure on health services require novel methodologies of assessment incorporating the preferences from groups of stakeholders across a set of multiple value dimensions. In this study an integrated multi-criteria approach was described simulating an HTA context for the case of advanced colorectal cancer treatments. Innovative approaches to decision-making for pricing and reimbursement of new therapies will be essential in the coming era of precision medicine and expensive but effective immunotherapies for cancer.

Future research could help validating the robustness of the methodology by conducting similar case studies with multi-stakeholder groups in different countries.

Chapter 7 – Paper 5

Application of the MCDA Value Framework: a Case Study on Metastatic Prostate Cancer with the Swedish Dental and Pharmaceutical Benefits Agency TLV¹²¹

Summary

Escalating drug prices have catalysed the generation of numerous “value frameworks” with the aim to inform payers, clinicians and patients around the assessment process of new medicines for the purpose of coverage and treatment selection decisions. Although this is an important step towards a more inclusive value-based assessment approach, aspects of these frameworks are based on weak and atheoretical methodologies and could potentially result in misleading recommendations or decisions.

Previous research has identified Multiple Criteria Decision Analysis (MCDA) as a potential methodology for assessing the value of drugs in Health Technology Assessment (HTA) and assist payers with resource allocation, however there are limited empirical evidence. The study applies in practice a theory-based MCDA framework through a proof-of-concept case study while engaging HTA decision-makers in a real world application.

A multi-attribute value theory methodological process involving problem structuring, model building, model assessment and model appraisal phases was adopted. The MACBETH approach was used: the different interventions were scored against the criteria through the development of value functions, weights were assigned to the criteria using a swing weighting technique and overall weighted preference value (WPV) scores were produced using simple additive aggregation. A facilitated decision analysis modelling approach was adopted as part of a decision conference.

The scope of the Swedish Dental and Pharmaceutical Benefits Agency (TLV) was used, but in addition supplementary evidence was considered for value concerns not explicitly addressed by TLV. Third-line biological treatments for mCRPC were evaluated, including cetuximab, panitumumab and aflibercept in combination with FOLFIRI, in

¹²¹ Chapter 7 has been submitted for peer review and it is single-authored by myself as: Angelis A. Applying Multiple Criteria Decision Analysis (MCDA) in the context of Health Technology Assessment (HTA): a case study on metastatic prostate cancer with the Swedish HTA agency TLV (Under Review at International Journal of Technology Assessment in Health Care)

metastatic castrate resistant prostate cancer (mCRPC) patients having received prior docetaxel chemotherapy in second line treatment. Eighteen attributes were initially considered, spanning value domains relating to therapeutic impact, safety profile, innovation level and socioeconomic impact.

A total of eight attributes were finally included. The therapeutic impact criteria cluster produced a relative weight of 44%, the safety profile cluster 33%, the socioeconomic impact cluster 15% and the innovation level cluster 7%. Enzalutamide scored the highest overall WPV score of 58.7, followed by abiraterone with 6.9 and cabazitaxel with 1.4 (partially because of a highly “negative” performance in the treatment discontinuation attribute). Dividing treatments’ purchasing costs with their overall WPV scores derived costs per unit of value at £419, £3,173, and £17,509 respectively.

Main challenges included the relatively subjective nature of setting the “lower” and “higher” reference levels on each attribute from single clinical studies given the lack of meta-analyses, the evaluation of clusters where participants had less experience, and the technical difficulties associated with ensuring that all evaluation criteria possess the ideal decision theory properties.

Overall, the participants of the workshop felt this was a useful exercise and that the methodological framework has the prospects of facilitating decision-making, mainly because it provides the opportunity to explicitly assess the performance of the options across a number of criteria, while eliciting trade-offs on their importance, explicitly and transparently. However additional research is recommended to address technical difficulties and enable its use as part of policy-making.

7.1 Background

In recent years, the assessment and appraisal of new and expensive medicines by Health Technology Assessment (HTA) bodies, health insurers, and gatekeeper agencies has received considerable attention, especially in countries with publicly funded health care systems. At times, this has emerged as a consequence of negative coverage decisions for some medicines, often related to treatments for high burden diseases associated with significant mortality and morbidity. As a consequence, intense disputes often arise usually involving patients, clinicians and pharmaceutical manufacturers on one hand and HTA bodies, commissioners of care and decision-makers responsible for the allocation of resources on the other.

To a large extent this conflict can be attributed to current methodologies used for the evaluation of new medicines as they cannot adequately capture the different notions of value evident given the existence of diverse perspectives ⁴²⁰. For example, despite the fact that the QALY metric has managed to successfully combine the two vital dimensions of health benefit relating to life expectancy and quality of life, its sole and inflexible adoption as part of economic evaluations in HTA can at times be regarded as blunt and insufficient not least because it does not adequately reflect other important value dimensions of the treatment ²¹⁰. This lack of comprehensiveness has led decision-makers to use additional parameters of value in an ad hoc and non-systematic manner often lacking transparency ³³⁹, with any inconsistency in the evaluation procedure threatening the perceived fairness of the decision and its acceptance from the public.

More recently a growing number of resource-conscious clinicians have started to publicly oppose the use of highly expensive new drugs with marginal incremental clinical benefit in clinical practice on the grounds of poor value for money. Escalating drug prices have catalysed the generation of numerous ‘value frameworks’ with the aim to inform treatment selection decisions by clinicians, possibly shared together with patients ^{170-172,174}. Although this is an important step towards a more inclusive value-based assessment approach, aspects of these frameworks are based on weak and atheoretical methodologies and could potentially result in misleading recommendations or decisions ¹⁷⁷.

Research and development of alternative methodologies for assessing the value of new medicines could overcome such limitations, contributing towards a more complete framework for measuring value and making resource allocation and treatment decisions. Multiple criteria decision analysis (MCDA) has recently surfaced as an alternative to traditional economic evaluation techniques with the prospects of addressing some of their limitations in HTA ^{72,90,91,191-194,332}, but also for eliciting patient preferences and facilitating treatment selection ^{340,452,453}.

This is a proof-of-concept case study utilising a new value framework the author recently developed for assessing the value of metastatic castrate resistant prostate cancer (mCRPC) treatment options following first/second line chemotherapy ⁴²¹. This work is a continuation of an earlier case study on the assessment of drugs for metastatic colorectal cancer involving preferences elicitation from a wide range of stakeholders in the English settings ⁴⁵⁴. It is the first from a series of case studies testing the value framework in practice by utilising value judgements and preferences from decision-makers of different national HTA agencies and social insurance bodies in Europe. A decision conference workshop ⁴⁴³

was organised with an experts panel consisting of assessors from the Swedish Dental and Pharmaceutical Benefits Agency (Tandvårds- och läkemedelsförmånsverket, TLV) that adopted a facilitated decision analysis modelling approach ^{94,422}. Metastatic prostate cancer was chosen as a case study topic mainly because of the disease severity and the availability of several expensive alternative treatment options highlighting a clear decision problem. Furthermore it has been a popular subject of appraisals by numerous HTA agencies, including TLV in Sweden and the National Institute for Health and Care Excellence (NICE) in England.

The adaptation and application of the value framework is described in the methods section, the preference value rankings of the treatment options are shown in the results section, with some value-for-money considerations and limitations of the methodology as a decision-making tool described in the discussion section.

7.2 Methods

7.2.1 Methodological Framework

An MCDA methodological process was adopted based on Multi-Attribute Value Theory (MAVT ^{54,205} involving the following five phases: (i) problem structuring, (ii) model building, (iii) model assessment, (iv) model appraisal, and (v) development of action plans ³³². The process was operationalised using a decision support system enabling the use of visual graphics to build a model of values, acting as a facilitation and decision-making tool to inform both the structuring phases (i, ii) and the evaluation phases of the process ⁴²⁹ (iii, iv). Additional information on the methodological framework can be found in *Appendix (7.1)*.

7.2.2 Clinical Practice and Scope of the Exercise (Problem Structuring)

Surgical or chemical castration are hallmarks of prostate cancer treatment since the 1940s ⁴⁵⁵. For many years prostate cancer's systemic therapy involved surgical or chemical continuous androgen deprivation ^{456,457}, often in combination with first-generation anti-androgens (i.e. combined androgen blockade) for patients who progress despite androgen deprivation and whose disease becomes hormone relapsed known as castrate resistant prostate cancer (CRPC) ⁴⁵⁸.

In 2004 clinical evidence demonstrated the survival benefit of docetaxel chemotherapy in combination with prednisolone for CRPC patients, when compared to

mitoxantrone in combination with prednisolone ^{459,460}. Since then a fruitful period in the development of prostate cancer drugs followed with a number of new molecules indicating significant clinical benefit in patients previously treated with docetaxel including abiraterone, cabazitaxel and enzalutamide ⁴⁶¹⁻⁴⁶³. Cross-resistance appears to exist between abiraterone and enzalutamide indicating that it is unlikely to experience clinical benefit by switching from one treatment to the other ^{464,465}, based on which NICE recommends only the use of either treatment and not both ^{466,467}. Although specific patient sub-populations might be contra-indicated to any of the treatments, in most cases all three new and expensive drugs could be suitable for patients and therefore the topic of CRPC treatment coverage acts as a suitable decision context for the application of the methodological framework in assessing their value.

Recent clinical studies have demonstrated a survival benefit for abiraterone or enzalutamide in CRPC patients in earlier line of treatment prior the administration of docetaxel chemotherapy ^{468,469}. However the expected number of patients receiving abiraterone or enzalutamide before docetaxel is likely to be shrunk following the publication of more recent clinical evidence supporting that docetaxel in combination with androgen deprivation therapy lead to high gains in survival of CRPC patients ^{470,471}. Consequently, the assessment of the three treatments - at the post-docetaxel stage - remains clinically relevant.

The analysis presented here focuses on assessing the value of third line treatments for mPC following prior docetaxel-containing (i.e. second line) chemotherapy, essentially adopting ESMO's clinical practice guidelines and the scopes of the respective Technology Appraisals (TAs) for each technology by NICE and TLV. In doing so, the same clinical evidence from the corresponding TAs was used to inform the respective criteria of the value tree, but this was further supplemented by additional evidence for value concerns not explicitly addressed by NICE and TLV. The scope of NICE TA255 was adopted for the case of cabazitaxel in combination with prednisone, the scopes of NICE TA259 and TLV TA4774/2014 were adopted for the case of abiraterone in combination with prednisone, whereas the scopes of NICE TA316 and TLV TA2775/2013 were adopted for the case of enzalutamide ^{466,467,472-474}.

7.2.3 Adaptation of the Advance Value Tree for Metastatic Prostate Cancer (Model Building)

Selection of evaluation criteria benefitted from a hybrid approach, which essentially combined features from “value focused thinking”³ and “alternative focused thinking”²⁰⁴. The author initially adopted a top-down approach for the selection of a generic set of evaluation criteria, prior the identification of any drug options, followed by a bottom-up approach for the adaptation of the generic criteria, which took place after the comparison of specific drug options’ characteristics³³².

More precisely, initially as part of the Advance-HTA project¹²² and following a systematic review of the literature and consultations with experts³³⁹, the author developed a generic value tree with higher-level criteria followed by lower-level criteria for assessing the value of new medicines in the HTA context⁴²¹. The structure of this emerging value tree (Advance Value Tree) consists of five value domains or criteria clusters, aiming to reflect all the essential value attributes of new medicines and their indication in a prescriptive manner. These included a) burden of disease (BoD), b) therapeutic impact (THE), c) safety profile (SAF), d) innovation level (INN) and, e) socioeconomic impact (SOC), with the first value domain relating to the disease of interest and the four latter domains relating to the actual medical intervention(s), essentially producing the following value function:

$$Value = f(BoD, THE, SAF, INN, SOC) \quad (6)$$

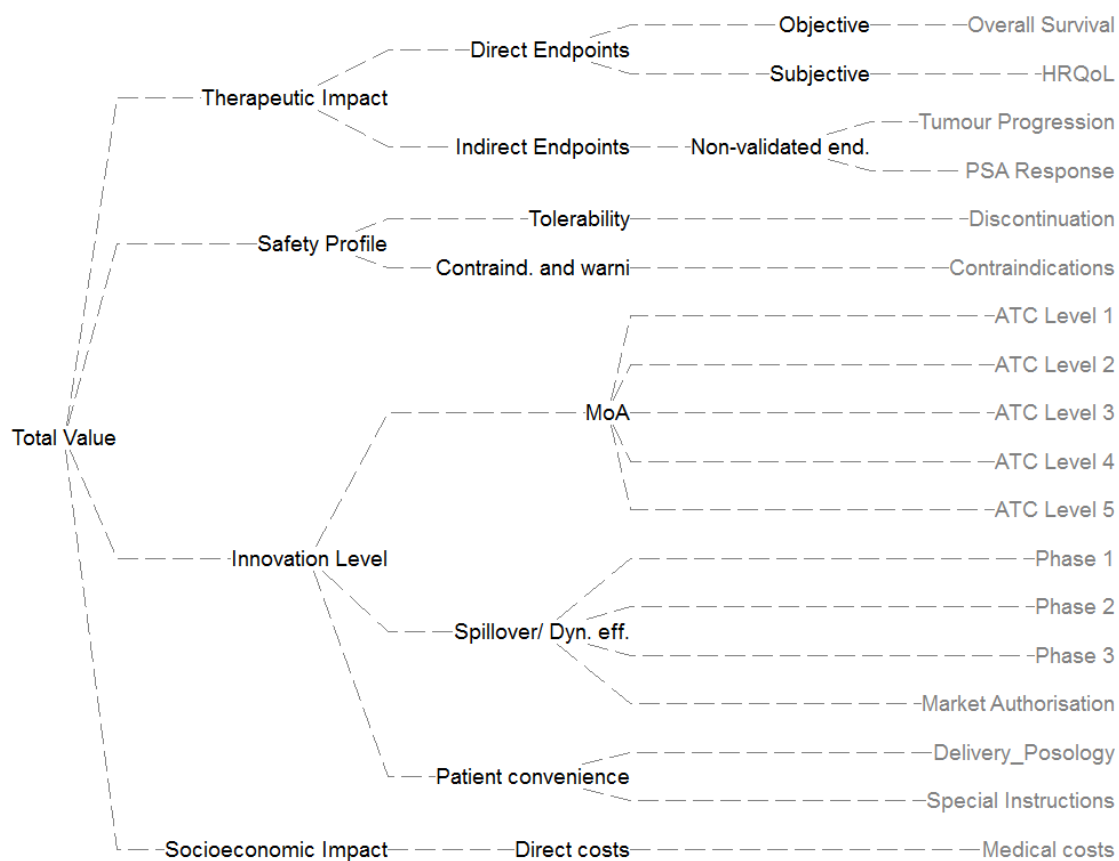
The generic value tree was adapted for the case of mCRPC while striving to adhere to the ideal decision theory properties to ensure methodological robustness²⁰⁶ producing a disease specific value tree with five criteria clusters decomposed into eight sub-criteria clusters and a total of 18 emerging criteria attributes (Figure 7.1). The definitions of all value attributes are listed in Table 7.1.

The value tree was validated later on with the decision conference participants, therefore reflecting another type of mixed approach adopted in terms of values construction: initially, model adaptation and structuring took place based on the literature and in

¹²² Advance-HTA is a research project funded by the European Commission's Seventh Research Framework Programme (FP7). It comprises several complementary streams of research that aim to advance and strengthen the methodological tools and practices relating to the application and implementation of Health Technology Assessment (HTA). It is a partnership of 13 Consortium members led by the London School of Economics - LSE Health.

consultation with a clinical expert, which was then collectively validated and evaluated with the group before the completion of the model building phase ⁴⁷⁵.

Figure 7.1: Preliminary value tree for metastatic prostate cancer (pre-workshop)



*image produced using the Hiview software version 3.2.0.4

Table 7.1: Attributes definition and sources of evidence

<i>Cluster</i>	<i>Attribute</i>	<i>Definition</i>	<i>Evidence source</i>		
			Abiraterone	Cabazitaxel	Enzalutamide
THERAPEUTIC IMPACT	Overall survival	The median time from treatment randomisation to death	de Bono et al 2011 ⁴⁶²	de Bono et al 2010 ⁴⁶¹	Scher et al 2012 ⁴⁶³
	HRQoL	Health related quality of life using EQ-5D score	TA259 ⁴⁶⁶ ; TA316 ⁴⁶⁷	N/A	TA316
	Radiographic tumour progression	The median survival time on which patients have not experienced disease progression (using RECIST criteria)	de Bono et al 2011	de Bono et al 2010	Scher et al 2012
	PSA response	The proportion of patients having a $\geq 50\%$ reduction in PSA	Fizazi et al 2012 ⁴⁷⁶	de Bono et al 2010	Scher et al 2012
SAFETY PROFILE	Treatment discontinuation	The proportion of patients discontinuing treatment due to AEs	de Bono et al 2011	de Bono et al 2010	Scher et al 2012
	Contra-indications	The existence of any type of contra-indication accompanying the treatment	EPAR ⁴⁷⁷ , Prescribing info	EPAR ⁴⁷⁸ , Prescribing info	EPAR ⁴⁷⁹ , Prescribing info
INNOVATION LEVEL	ATC Level 1	The technology's relative market entrance in regards to its ATC Level 1 (Anatomical)	WHO ATC index ⁴³⁹	WHO ATC index	WHO ATC index
	ATC Level 2	The technology's relative market entrance in regards to its ATC Level 2 (Therapeutic)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 3	The technology's relative market entrance in regards to its ATC Level 3 (Pharmacological)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 4	The technology's relative market entrance in regards to its ATC Level 4 (Chemical)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 5	The technology's relative market entrance in regards to its ATC Level 5 (Molecular)	WHO ATC index	WHO ATC index	WHO ATC index

	Phase 1	The number of new indications for which the technology is investigated in Phase 1 clinical trials	ClinicalTrials.gov ⁴⁴⁰	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 2	The number of new indications for which the technology is investigated in Phase 2 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 3	The number of new indications for which the technology is investigated in Phase 2 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Marketing authorisation	The number of new indications that the technology has gained an approval for at the stage of marketing authorisation	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Delivery Posology	The combination of the delivery system (RoA and dosage form) with the posology (frequency of dosing and duration of administration) of the treatment	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
	Special instructions	The existence of any special instructions accompanying the administration of the treatment	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
SOCIO-ECONOMIC IMPACT	Medical costs impact	The impact of the technology on direct medical costs excluding the purchasing costs of the technology	BNF 69, Prescribing info, Connock et al 2011 ⁴⁸⁰ , Riemsma et al 2013 ⁴⁸¹ , TA259	BNF 69, Prescribing info, de Bono et al 2010, TA255 ⁴⁷²	BNF 69, TA316

7.2.4 Evidence Considered and Alternative Treatments Compared (Model Building)

The alternative treatment options compared in the analysis include cabazitaxel (Jevtana ®) in combination with prednisone, abiraterone (Zytiga ®) in combination with prednisone and enzalutamide (Xtandi ®) monotherapy.

The evidence sources used include the peer review publications concerning the pivotal clinical trials of the alternative treatment options that were considered in the appraisals by NICE and TLV ^{461-463,476}, NICE Evidence Review Group (ERG) reports ^{480,481}, or peer review studies coming out of them ^{482,483} and a Swedish population study on health related quality of life ⁴⁸⁴, Product Information sections of the European Public Assessment Reports (EPAR) from the European Medicines Agency (Annex I and III) ⁴⁷⁷⁻⁴⁷⁹, Anatomical Therapeutic Chemical (ATC) classification system indexes through the portal of the WHO Collaborating Centre for Drug Statistics Methodology ⁴³⁹, and ClinicalTrials.gov listings ⁴⁴⁰. The clinical evidence (falling under the THE and SAF criteria clusters) for each alternative treatment option were sourced from the same clinical studies that NICE and TLV evaluated (single pivotal trials for each drug). The source of evidence used for identifying the performance of options across the evaluation criteria is shown in Table 7.1. Additional information on the evidence considered can be found in *Appendix (7.2)*.

7.2.5 Setting Attributes' Ranges and References Levels (Model Building)

In regards to model building, the attribute ranges were selected so that they would be bound by the minimum (min) and maximum (max) attribute limits. Within these min-max attribute limits, intermediate “lower” (x_l) and “higher” (x_h) reference levels were defined to serve as benchmarks for defining the preference value scores of 0 and 100 respectively, for constructing criteria value functions and eliciting their relative weights^{123 449}. The emerging value scores of options could therefore take negative or higher than 100 values where $v(x_{\text{lower}}) = 0$ and $v(x_{\text{higher}}) = 100$, essentially by conducting a linear transformation which is acceptable as an interval scale such as a value scale.

Limits were assigned that included the current performance of the options as well as possible short-term future changes, ensuring that the limits would still be considered realistic by the assessors, essentially reflecting a “local” value scale adjusted for future expected performance². Table 7.2 outlines “lower and higher reference levels” for all attributes at the

¹²³ these are interval scales and therefore is important to set up clear bounds/limits for each attribute

pre-workshop stage, and the basis of their selection¹²⁴. For the case of the clinical attributes (falling under the THE and SAF value domains), the reference levels were decided in consultation with a clinical expert (urologic oncologist).

The methodological basis adopted for setting the min-max attribute limits and the choice of the reference levels was the following. For the case of clinical therapeutic attributes, “lower reference levels” were based on BSC figures, coming from the median of the respective placebo arm of the *AFFIRM* trial, with the exception of the HRQoL attribute (EQ-5D index score) that was based on the utility of stable disease with no treatment coming from past NICE TAs^{466,472}. The “higher reference levels” were derived by adding a 20% absolute increment to the performance level of the best performing option, besides for the case of the HRQoL attribute (EQ-5D index score) that was based on the general Swedish population⁴⁸⁴. The rationale was to use a “satisfactory performance” (proxied by BSC) and an “ideal performance” which could offer a flexibility margin to incorporate the performance of future improved options (proxied by 20% higher than today’s available), corresponding to the 0 and 100 anchor levels of the value function scale respectively, with options performing better than the satisfactory level scoring more than 0. Consequently two reference levels within the attribute range were defined in most cases: i) the “lower reference level” (x_l) (i.e. BSC-based satisfactory performance), acting on the same time also as the minimum limit of the attribute range (x_*); and ii) the “higher reference level” (x_h) (i.e. 20% higher than the best performing option), acting on the same time as the maximum limit of the attribute range (x^*) to give $x_* = x_l < x_h = x^*$.

Similar but reverse logic was adopted for setting the reference levels in the “treatment discontinuation” attribute of the safety cluster; the “lower reference level” was defined equal to the BSC (i.e. placebo) arm of the *AFFIRM* trial. However, contrary to the logic adopted so far for the therapeutic impact criteria, the “higher reference level” was not set equal to 20% less (because the lower the figure the higher the value) than the best performing option, but rather equal to the minimum natural limit of the attribute scale (i.e. 0%) which was regarded as an “ideal” level. In turn, the minimum limit of the scale was derived by incorporating 20% to the worst performing treatment option. A similar approach was used for setting the reference levels of the qualitative attribute of “contraindications”, defining the “higher reference level” equal to the maximum (i.e. most attractive) limit of the attribute scale (i.e. no contraindications) and the “lower reference level” equal to the minimum (i.e.

¹²⁴ assuming no impact of LHRH analogue

least attractive) limit of the attribute scale, based on the performance of the alternative options therefore acting as reference levels of a “local” scale.

For the innovation attributes, the “higher reference level” was set either equal to 20% higher than the best performing option for the case of natural quantitative attributes (e.g. number of new indications for which the technology is investigated in a given clinical development stage), or equal to the maximum limit of the scale for the case of constructed qualitative attributes (e.g. the existence of any special instructions, the technology's relative market entrance in regards to its ATC Level. Given that the BSC performance was irrelevant to be used as satisfactory level in the innovation attributes, and the fact that any efforts to derive a “satisfactory” level would by definition be subjective in nature, the minimum limit of the scale for each attribute was used as a “lower reference level”. Therefore the “lower reference level” was based on the worst performance plausible as inferred from the lowest limit of the scales, both for the case of natural quantitative attributes (e.g. 0 number of new indications for which the technology is investigated in a given clinical development stage), and the case of constructed qualitative attributes (e.g. worst possible combination of special instructions, 5th entrance at an ATC level).

For the socioeconomics attribute (impact on direct costs), the “higher reference level” was based on the BSC’s impact on cost (i.e. £0 impact on costs), given that by definition impact on costs for all treatment options are incremental to BSC, and the “lower reference level” was derived by adding a 20% absolute increment to the worst performing option (i.e. to the one with the biggest impact on costs).

For the purpose of eliciting preferences and producing the matrix of judgments using M-MACBETH, two additional intermediate attribute levels were aimed to be incorporated between the two defined attribute levels (giving a total of four different attribute levels) so that the granularity of the scale is increased, essentially to improve the representation of any differences in value across the attribute ranges. In cases where the gaps between the two defined levels were disproportionate larger, a third intermediate level was added for a more homogeneous dispersion, giving a total of five attribute levels (two defined and three intermediates), whereas in cases of disproportionate smaller gaps, one intermediate level was added giving a total of three attribute levels (two defined and one intermediate). In no cases there were less than three and more than five attribute levels in total, except for the case of “new indications at MA stage” for which the “higher reference level” was equal to 1 and the “lower reference level equal to 0”.

7.2.6 Decision Conference (*Model Assessment and Appraisal*)

The model assessment and model appraisal phases of the exercise took place through a facilitated workshop with assessors from TLV, taking place as a decision conference and hosted at the head offices of TLV in Stockholm, Sweden. In total four experts acted as participants, including one medical investigator, two health economists and one chief pharmacist. Background material introducing the scope of the exercise in more detail was sent to the participants one week before the workshop.

The author acted as an impartial facilitator and assisted the group's interaction and thinking about the decision problem using the preliminary version of the mCRPC specific value tree (Figure 7.1) and the relevant data as the model's starting point, based on which value judgements and preferences were elicited on-the-spot ^{416,422,444}. On the day of the workshop the preliminary model was validated with the participants by revising it one cluster by cluster in real time through an open discussion, seeking group consensus (i.e. majority agreement if not full agreement) by adopting an iterative and interactive-model-building process where debate was encouraged and differences of opinion actively sought.

In terms of the decision-aiding methodology used, the author acted as an impartial facilitator with the aim of enhancing content and process interaction, while refraining from contributing to the content of the group's discussions, essentially guiding the group in how to think about the issues but not what to think ^{416,446}. In terms of facilities, the room of the workshop had a II-shaped meeting table for all the participants to have direct eye contact, with an overhead projector screen surrounded by whiteboards. The M-MACBETH software was operated using a laptop, the screen of which was connected to the projector.

The workshop lasted two half-days (2 sessions in total), three to four hours each session, with a short coffee break around the middle of each session. In the first day, the workshop started with an overview of the MCDA methodology adopted and the description of the value tree. The value tree was then presented and analysed cluster by cluster. At the beginning of each cluster the value tree was validated; the various criteria were explained, followed by a group discussion relating to their relevance and completeness. As a result of this iterative process, some of the criteria were removed because they were perceived as irrelevant or non-fundamental. Then, value functions were elicited for the different criteria and relative weights were assigned within the clusters. Finally, relative weights were assigned across clusters, enabling the calculation of the options' overall WPV scores.

Table 7.2: Pre-workshop attribute reference levels and basis of selection

<i>Cluster</i>	<i>Attribute name</i>	<i>Attribute metric</i>	<i>Lower level</i>	<i>Basis</i>	<i>Higher level</i>	<i>Basis</i>
THERAPEUTIC IMPACT	Overall survival	months	13.6	BSC	22.1	20% higher than the best performing option
	HRQoL	utility (EQ-5D)	0.72	Utility used for progressive disease in TA259	0.82	Utility scores of general population
	Radiographic tumour progression	months	2.9	BSC	10.6	20% higher than the best performing option
	PSA response	% patients	1.5	BSC	64.8	20% higher than the best performing option
SAFETY PROFILE	Treatment discontinuation (% of patients)	% patients	10	BSC	0	Maximum limit of the scale
	Contra-indications	types of contra-indications	Hypersensitivity + hepatic impairment + low neutrophil counts	Minimum limit of the scale	No contra-indications	Maximum limit of the scale
INNOVATION LEVEL	ATC Level 1	relative market entrance	5	Minimum limit of the scale	1	Maximum limit of the scale
	ATC Level 2	relative market entrance	5	Minimum limit of the scale	1	Maximum limit of the scale
	ATC Level 3	relative market entrance	5	Minimum limit of the scale	1	Maximum limit of the scale

	ATC Level 4	relative market entrance		5	Minimum limit of the scale		1	Maximum limit of the scale
	ATC Level 5	relative market entrance		5	Minimum limit of the scale		1	Maximum limit of the scale
	Phase 1	number of new indications		0	Minimum limit of the scale		10	20% higher than the best performing option
	Phase 2	number of new indications		0	Minimum limit of the scale		16	20% higher than the best performing option
	Phase 3	number of new indications		0	Minimum limit of the scale		2	20% higher than the best performing option
	Marketing authorisation	number of new indications		0	Minimum limit of the scale		1	20% higher than the best performing option
	Delivery Posology	types of delivery system & posology combinations	Oral, every day - one off + IV, every 3 weeks - 1 hour*		Minimum limit of the scale	Oral, every day - one off*		Maximum limit of the scale
	Special instructions	types of special instructions	No food + concomitant and/or pre-medication*		Minimum limit of the scale	None*		Maximum limit of the scale
SOCIO-ECONOMIC IMPACT	Medical costs impact	GBP (£)		10,000	20% higher than the worst performing opt.		0	BSC

7.2.7 MCDA Technique (Model Assessment and Appraisal)

A value framework based on a value measurement method was adopted making use of a simple additive aggregation model (i.e. linear, weighted average approach) which assumes preference independence between the different attributes, with overall value $V(.)$ of an option a defined by Equation 4⁴²¹. Overall, the additive value model is assumed as a working hypothesis and therefore the model construction is done to respect its underlying properties (see section 1.1.4), requiring positive weights summing one and the use of explicit reference values of 0 and 100, see Table 7.2.

Value functions were elicited from the workshop participants using the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH), a pairwise qualitative comparison approach where qualitative judgements about the difference of value between different pairs of attribute levels are expressed in word categories (no difference, very weak, weak, moderate, strong, very strong, extreme, or anything in between)^{406,447}. By only requiring semantic judgements which are then converted into a cardinal scale, MACBETH provides a simple, interactive and constructive approach, with various real world applications illustrating its usefulness as a decision support tool⁴⁰⁸⁻⁴¹¹, while being based on strong theoretical foundations⁴⁰⁷. An indirect (qualitative) swing weighting technique was used to elicit relative criteria weights⁴⁴⁷, given that direct questions of importance for a criterion that do not take into account their attribute ranges are known to be one of the most common mistakes in making value trade-offs²⁰².

In turn, M-MACBETH was used as a decision support system in order to construct the value tree, elicit criteria value functions (based on which options were scored), assign relative criteria weights through a qualitative swing weighting protocol, combine preference value scores and weights together using additive aggregation (i.e. simple additive model) to derive overall weighted preference value (WPV) scores, and perform sensitivity analysis on criteria weights^{448,449}. The software automatically performs consistency checking between the qualitative judgements expressed, and in addition a second consistency check was manually performed by the author to validate the cardinality, i.e. interval type, of the emerging value scale by comparing the sizes of the intervals between the proposed scores and inviting participants to adjust them if necessary⁴⁵⁰, a requirement which is essential for the application of simple additive value models. More information regarding the technical details of MACBETH is available in *Appendix (7.3)*.

7.2.8 Costs Calculation (*Model Assessment and Appraisal*)

Drug costs were calculated according to UK prices (excl. VAT), pack sizes and dosage strengths as found on the British National Formulary (BNF 68); the recommended dosages and treatment durations were taken from its pivotal trial for the case of cabazitaxel ⁴⁶¹, its' respective NICE technology appraisal for the case of enzalutamide ⁴⁶⁷, and its labelling and package leaflet (EPAR- Annex III) for the case of abiraterone ⁴⁷⁷. Vial wastage was assumed in all calculations. Drug administration costs for cabazitaxel were kept consistent with the respective NICE TA⁴⁷². The UK perspective was adopted as a neutral benchmark partly because of the readily available data but primarily to allow the measurement of cost(s) in a common unit across a series of similar case studies in different European countries, so that overall WPV scores can then be viewed against the same cost denominator to produce directly comparable cost-value ratios.

7.3 Results

7.3.1 Criteria Validation and Amended Value Tree for Metastatic Prostate Cancer

Overall, consensus (i.e. full agreement) was reached relatively easily in terms of criteria inclusion and exclusion. Most importantly though, no criteria were deemed to be missing. The final version of the value tree, as emerged following the open discussion with the participants of the workshop is shown in Figure 7.2. In total, 10 out of the 18 attributes were removed from the value tree because they were judged from the participants to be non-fundamental for the scope of the exercise. Most of these attributes however were lying under the innovation level cluster, with all attributes relating to ATC level of the drugs (five attributes) and spill-over effects (four attributes) being eliminated. The only additional attribute removed was the PSA response attribute under the therapeutic impact cluster, because it deemed to reflect no value of concern given the existence of the remaining therapeutic attributes.

An example of a value judgements matrix and its conversion into a value function for the case of the Overall Survival (OS) is shown in *Appendix (7.4)*. Effectively all attributes produced linear value scales, either based on quantitative or qualitative performance levels. The value functions for all attributes are shown in *Appendix (7.4)*.

7.3.2 Performances of Options, Criteria Weights and Overall Preference Value Rankings

The performance of the options across the different attributes together with the “lower” and “higher” reference levels is shown in Table 7.3.

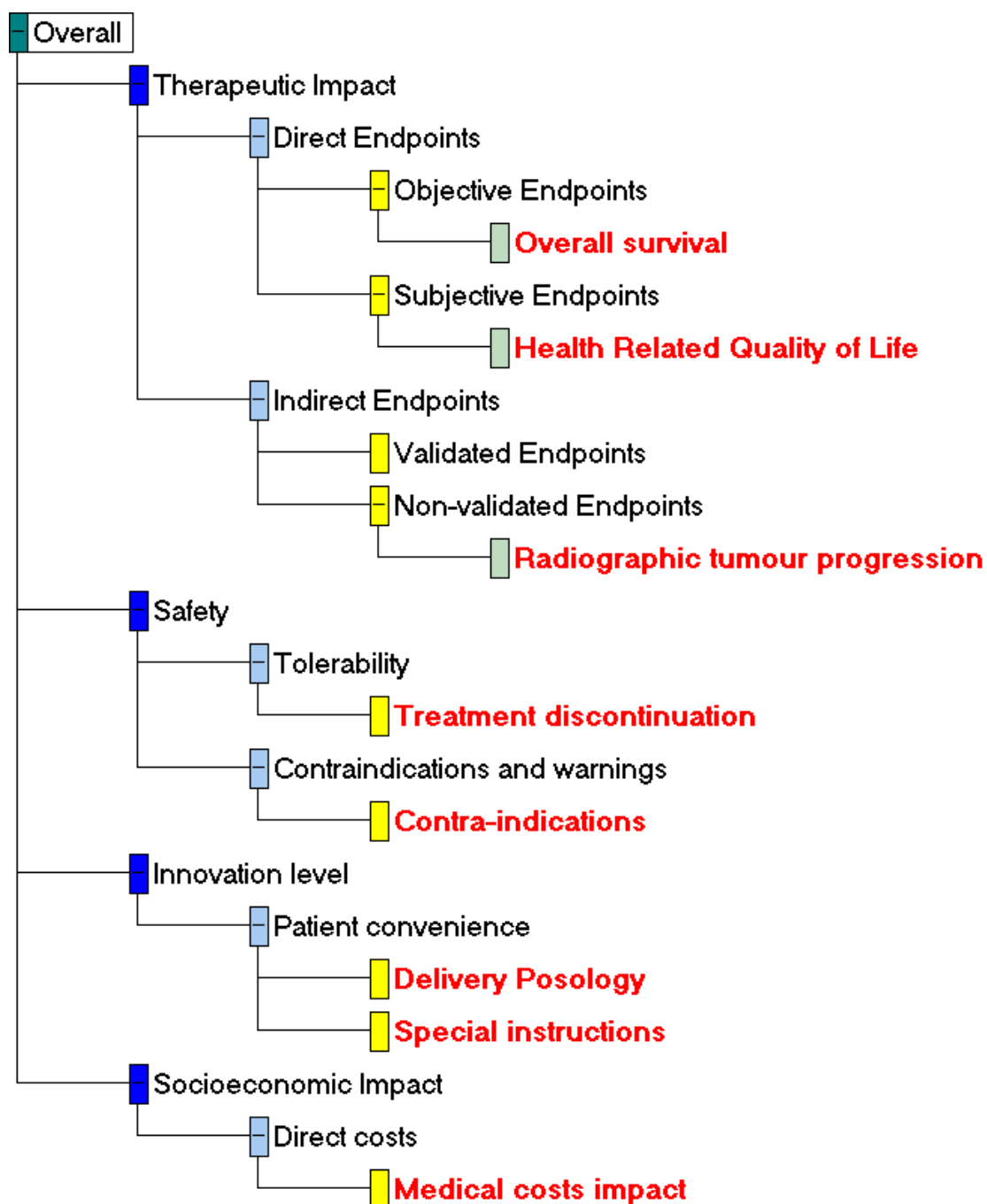
The overall WPV scores of the options and their break down into their partial value scores across the different criteria attributes with their respective weights are shown in Table 7.4. Enzalutamide scored the highest overall WPV score of 58.7. Abiraterone and cabazitaxel produced overall WPV scores of 6.9 and 1.4 respectively, partially because of relatively high “negative” performances in the treatment discontinuation attribute, producing absolute preference value scores of -95.3 and -87.5 respectively, and weighted preference value scores of -20.2 and -18.6 respectively. This was due to the fact that their performance on the treatment discontinuation attribute (19% and 18% respectively), lied below the lower reference level of the scale (i.e. 10%). A stacked bar plot of the overall WPV scores of the alternative treatments across the attributes is shown in Figure 7.3.

The relative weights assigned to the different attributes are shown in Figure 7.4. By taking into account the “lower”-“higher” ranges of the attributes, the greatest relative weights were yielded for the case of Overall Survival, Treatment Discontinuation and HRQoL (with relative magnitudes of 23, 21, and 15 out of 100 respectively), adding up a combined relative weight of 60% of the total. In regards to the total weights assigned across the criteria clusters, therapeutic impact cluster totaled a relative weight of 44 in total (three attributes), the safety profile cluster totaled a relative weight of 33 (two attributes), the innovation level cluster totaled a relative weight of 7 (two attributes), and the socioeconomic impact cluster totaled a relative weight of 15 (single attribute) out of 100.

7.3.3 Value for Money Analysis

A plot of the options’ overall WPV scores versus their costs (purchasing costs plus any administration costs) is shown in Figure 7.5. By using rounded up cost figures of £24,600 for enzalutamide, £21,900 for abiraterone and £23,900 for cabazitaxel (£22,190 drug cost and £ 1,710 administration cost) and dividing them with overall WPV scores, the costs per unit of MCDA value were calculated to be £419, £3,173, and £17,509 respectively (Table 7.4). Therefore in terms of value-for-money, cabazitaxel is shown to being dominated by abiraterone, while also being very close to be dominated by enzalutamide (£500 difference). Enzalutamide on the other hand is associated with a higher cost (£2,500 difference) and a higher overall WPV score (51.8 difference) compared to abiraterone.

Figure 7.2: Final value tree for metastatic prostate cancer (post-workshop)



* image produced using the M-MACBETH (beta) software version 3.0.0

7.3.4 Sensitivity and Robustness Analysis

Parameter uncertainty relating to the estimation of weights was addressed by conducting deterministic sensitivity analysis. At the end of the workshop, the impact of baseline weights' changes on options' rankings was explored. In order for cabazitaxel to become better ranked than abiraterone, the PFS relative weight would have to change from 5.8 to 15.6, the Treatment Discontinuation relative weight from 21.2 to 53.9, or the Special Instructions relative weight would have to increase from 4.1 to 12.2 (figures shown in *Appendix, 7.5*). Alternatively, for cabazitaxel to become better ranked than enzalutamide, the PFS relative weight would have to change from 5.8 to more than 89.5 (figure shown in *Appendix, 7.5*). No scenario would result in abiraterone being ranked above enzalutamide. Therefore, conclusions were robust as the ranking of the treatments was not sensitive to single variations of up to at least 100% along the attributes' weight range. The robustness of the results was also validated by conducting 8-way sensitivity analysis in the reference levels of the attributes using the respective function of the M-MACBETH software ("Robustness analysis"), which showed that a simultaneous change of up to 10% across all of the attribute reference levels would not impact the ranking of the alternative treatments (figure shown in *Appendix, 7.5*).

Other types of uncertainty also exist, including stochastic uncertainty, structural uncertainty and heterogeneity that could possibly be addressed with other approaches (or combinations of techniques), such as probabilistic sensitivity analyses, Bayesian frameworks, fuzzy set theory or grey theory²⁰⁸. For instance, if significant uncertainty exists with regards to option performance due to sampling variation from clinical trials, or in terms of criteria weights due to lack of agreement with them, the application of point estimates might be inappropriate in which case stochastic multi-criteria acceptability analysis (SMAA) could be used⁴⁵¹.

Table 7.3: Options performance across the criteria attributes

Attribute name	Attribute metric	Lower level	Abiraterone	Cabazitaxel	Enzalutamide	Higher level
Overall survival	months	13.6	15.8	15.1	18.4	22.1
HRQoL	utility (EQ-5D)	0.72	0.76	0.76	0.76	0.82
Radiographic tumour progression	months	2.9	5.6	8.8	8.3	10.6
Treatment discontinuation	% of patients	10	19	18	8	0
Contra-indications	types of contra-indications	hyp + hep imp + low neut	hyp + hep imp	hyp + hep imp + low neut	hyp	None
Delivery Posology	types of delivery system & posology combinations	oral, daily - one off + IV, every 3 wks - 1 hr	oral, daily - one off	oral, daily - one off + IV, every 3 wks - 1 hr	oral, daily - one off	oral, daily - one off
Special instructions	types of special instructions	Concomitant and/or pre- med + no food	Concomitant and/or pre- med + no food	Concomitant and/or pre- med	None	None
Medical costs impact	GBP (£)	10,000	5,750	7,992	567	0

*Used the same score of the other two options as data not available. hyp: hypersensitivity; hep imp: hepatic impairment; low neut: low neutrophil count

Table 7.4: Overall weighted preference value (WPV) scores, partial preference value scores, relative weights, costs and cost per unit of value

	Lower level	Abiraterone	Cabazitaxel	Enzalutamide	Higher level	Relative Weights
Overall WPV score	0	6.9	1.4	58.7	100	100
Overall survival	0	26.2	17.9	56.3	100	23
HRQoL	0	40.0	40.0	40.0	100	15
Radiographic tumour progression	0	26.5	74.0	66.8	100	6
Treatment discontinuation	0	-95.3	-87.5	20.0	100	21
Contra-indications	0	33.3	0.0	83.3	100	12
Delivery Posology	0	100.0	0.0	100.0	100	3
Special instructions	0	0.0	60.0	100.0	100	4
Medical costs impact	0	40.5	19.1	93.7	100	15
<i>Cost (£)</i>		21,900	23,900	24,600		
<i>Cost per unit of value</i>		3,173	17,509	419		

HRQoL: health related quality of life

Figure 7.3: Stacked bar plot of treatments’ overall weighted preference value (WPV) scores across all attributes

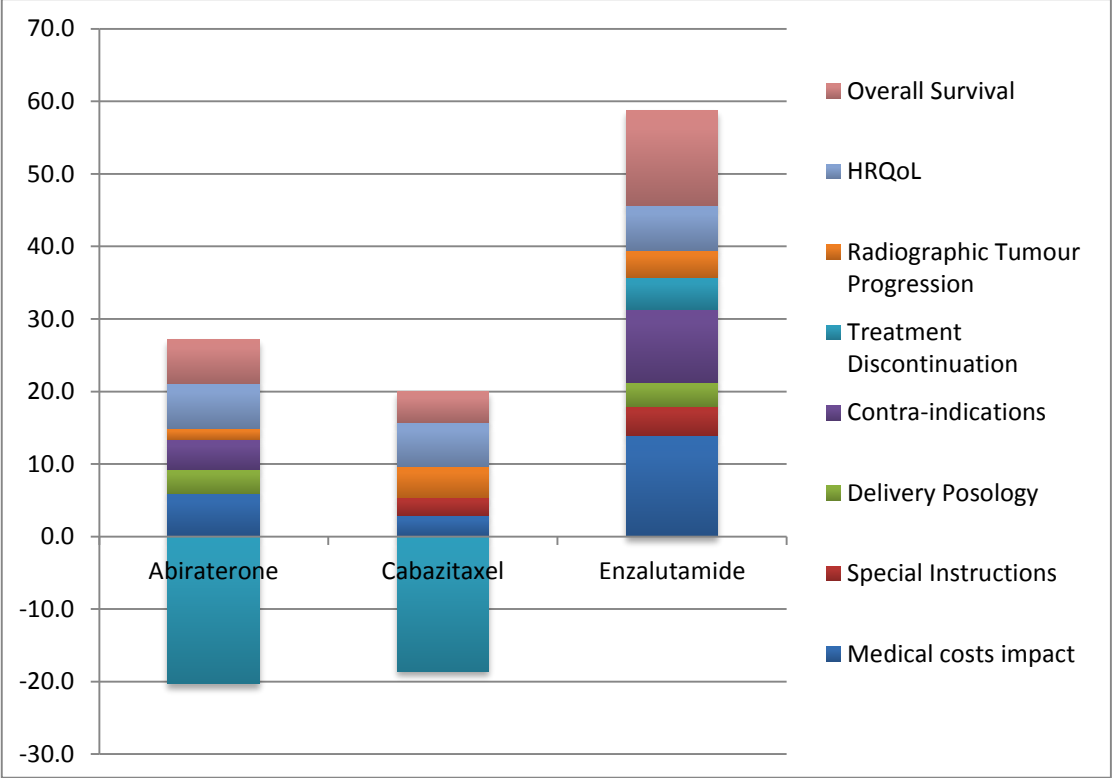


Figure 7.4: Criteria weights stacked bar

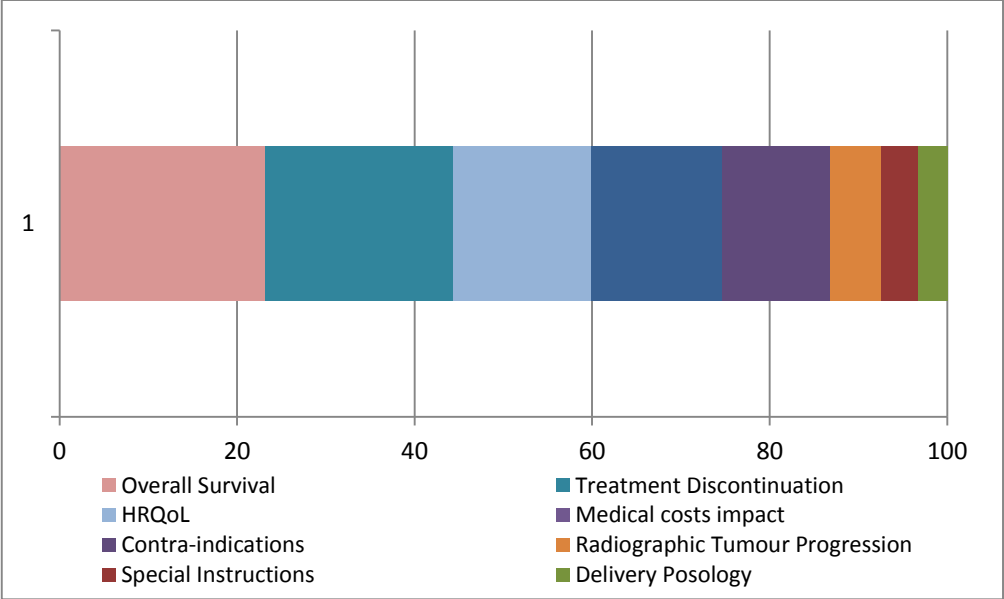
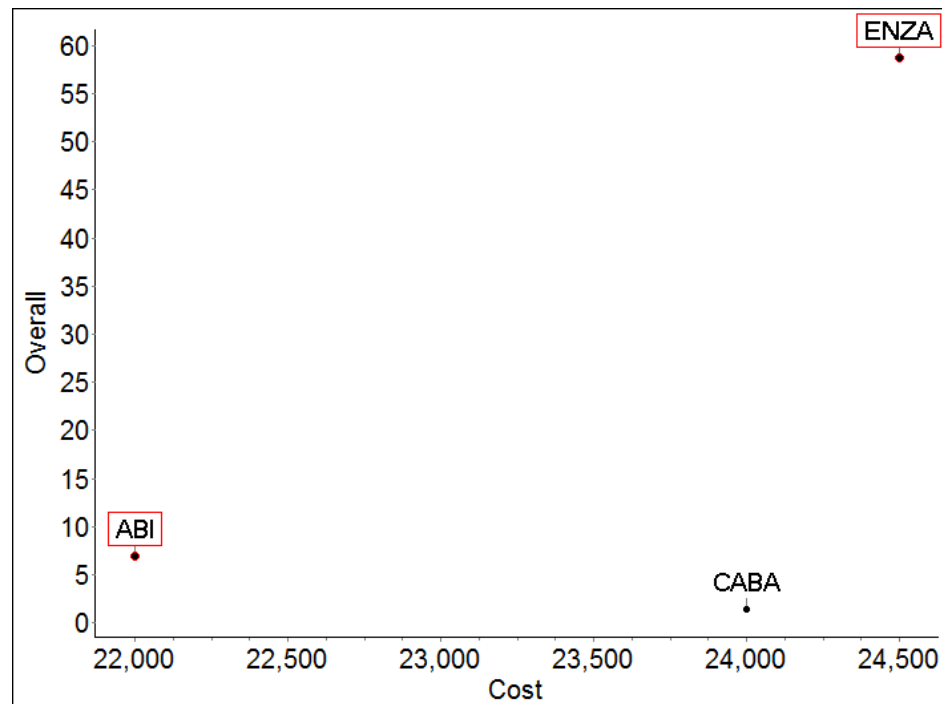


Figure 7.5: Cost benefit plot of overall weighted preference value (WPV) scores versus costs



* image produced using the M-MACBETH (beta) software version 3.0.0

7.4 Discussion

This case study demonstrated an application of a recently developed MCDA methodological process and value framework in practice ^{332,421}, for a set of alternative mCRPC treatments from the perspective of an HTA agency. In terms of design, implementation and review of the analysis, the process adopted is effectively in alignment with the recent ISPOR good practice guidelines on the use of MCDA for health care decisions ²⁰⁰.

The participants of the workshop felt this was a useful simulation exercise and that the value framework has the prospects of acting as a valuable decision supporting tool, mainly because it provides the opportunity to explicitly assess the performance of a set of options across an encompassing number of evaluation criteria, while eliciting trade-offs on their relative importance, flexibly and transparently.

In regards to the removal of the spill-over effect attributes under the innovation level cluster, this took place on the grounds that they “currently go beyond the agency’s remit” and therefore should not be considered in the first place. However, the opinion was evident that in reality this information could act as supplementary evidence in

current practice, not primarily for affecting the coverage of the respective drugs, but possibly for other secondary purposes like awarding reimbursement extensions or communicating internally possible new indications in the near future. In any case, all innovation spill-over attributes were excluded.

7.4.1 Value for Money Considerations

Alternative options were assessed and ranked based on their overall WPV scores reflecting their performance against an explicit set of evaluation criteria weighted for their relative importance based on the preferences of the group, therefore acting as a more holistic benefit component. Incorporation of drug costs (purchasing and administration costs) enabled the estimation of “cost per unit of value” ratios which showing no overall “value for money” dominance between the treatments but only between two of them.

Budget impact considerations could be incorporated in the cost-value ratio estimates at system level by taking into account the expected number of patients receiving the treatments; however, given that the scale of costs is not restricted in contrast to benefits, costs could be overestimated compared to benefits which could produce a bias towards the cheapest options, and therefore costs and benefits should ideally be estimated in comparable terms such as a per-patient basis ²⁰⁰.

Future research might need to investigate the opportunity cost associated with disinvestments which could feed the development of incremental cost value ratio (ICVR) thresholds, acting as efficiency cut-off points similarly to current incremental cost effectiveness ratio (ICER) thresholds used in economic evaluations. This would face however all the current challenges associated with the technical difficulties in estimating sound ICER thresholds based on opportunity cost ^{485,486}.

7.4.2 Limitations and Challenges

Among the limitations of the case study was the use of un-synthesised evidence to inform the clinical attributes given the lack of relative treatment effects, therefore limiting the extent to which the results can inform real policy-making. An important challenge would relate to the technical difficulties in ensuring that all attributes possess the required theoretical properties, and more specifically the fact that potential preference-dependence was observed between the OS and HRQOL attributes which

was addressed by deriving OS value functions that were effectively conditional on the range of the HRQOL attribute. However, the use of conditional value functions might be incompatible with the use of an additive model and, therefore, it should be used with caution, as for example within descriptors of performance. In order to be able to use an additive model under the existence of preference dependence, the effect between the two attributes would have to equal zero, or in other words the product of their value scores would have to be zero; for a more extensive discussion around the use of linear additive model together with conditional value functions due to preference dependence see section 8.3.2. Another limitation would be that the HRQOL of the progressive disease was not assessed because none of the treatments assumed to have any effects on it, something which might not hold true in other conditions. Among the main challenges was the relative subjectivity in setting “higher” and “lower” reference levels, something which was addressed through expert opinion and investigated in secondary analysis following the workshop. These limitations and challenges are extensively discussed in the Limitations and Challenges section in the last chapter of the thesis (section 8.3.2).

7.5 Conclusion

In health care systems with significant budgetary pressures HTA challenges relating to the evaluation of and resource allocation for novel treatments require novel methodologies of assessment with a more encompassing value-based assessment approach. Such methodologies should be based on robust theoretical principles so the results can lead to credible decisions. In this case study a multi-criteria methodology was tested to evaluate expensive therapies for the treatment of metastatic prostate cancer while taking into account the preferences of assessors from TLV, producing a transparent ranking of their value across a number of value dimensions. Future research could help validate the robustness and usefulness of the current value framework by conducting similar case studies with other HTA bodies across different European countries.

Chapter 8 – Conclusions, Policy implications, Limitations and Future Research Directions

8.1 Key Conclusions

8.1.1 Identification of Limitations in Current Value Assessment Approaches of New Medicines

Current evaluation approaches in HTA examine the comparative clinical benefit or cost-effectiveness of new medicines as part of formal and explicit processes, through the elicitation of ‘Scientific Value Judgements’ (ScVJ) relating to their clinical performance and usually in relation to their costs, therefore addressing comparative efficacy/effectiveness and efficiency concerns. However, important inconsistencies and uncertainties exist on how other value concerns are accounted for, especially those relating to dimensions of benefit falling under the evidence domains of burden of disease, innovation level and socioeconomic impact.

Principal among the lessons learned, was the fact that although such ‘Social Value Judgements’ (SoVJ) might exert an influence on decisions, they rarely take place as part of a structured process; instead, they are often taken into account in an informal way or an *ad hoc* basis following non-transparent discussions. The selectivity in measuring the performance of new medicines across these social value dimensions as part of a deliberative process, in addition to the lacking clarity around their relative importance to the overall decision and the value trade-offs decision-makers are willing to make, both keep the interplay between SoVJ and ScVJ unknown. This can diminish the reasonableness, efficiency and credibility of the decision-making process and its outcomes.

Research findings pointed towards the need to develop transparent methodologies that allow the explicit incorporation and structured analysis of preferences around multiple criteria and their value trade-offs. This can lead to more rational evidence-based decision-making, which could improve the efficiency of resource allocation decisions and as a result increase societal welfare, while on the same time raising public confidence and the perceived fairness of decision outcomes.

8.1.2 Conceptual Verification of MCDA as a Viable Alternative Approach for HTA

The thesis proposes the use of MCDA as an alternative to economic evaluation in the context of HTA. It is argued however that the use of MCDA needs to adhere to decision analysis theory for the results to be robust so that they can facilitate decision-making. An MCDA methodology process based on Multi Attribute Value Theory (MAVT) for the context of HTA is recommended which can be divided into the phases of problem structuring, model building, model assessment, model appraisal, and action plans. A “value-alternative hybrid thinking” approach can be adopted for structuring a value tree in model-building, according to which a generic set of criteria are created first as part of a top-down approach driven by the overall value concern, followed with the adaptation of the criteria for the particular decision-making context and the definition of attributes as part of a bottom-up approach driven by the decision alternatives as part of an “alternative-focused thinking” approach. For the analysis to be robust and for decision recommendations to be ultimately meaningful, criteria and attributes should adhere to a number of properties which could act as best practice requirements for the model-building phase.

Implementation of MCDA methodologies and their linkage with policy-making could take place in the form a supplementary “incremental” mode to cost-effectiveness analysis (CEA) adjusting the incremental cost effectiveness ratio (ICER) through the incorporation of additional benefit dimensions, or in the form of a pure “clean-slate” mode where value is derived without the use of CEA.

Among the lessons learned from the adoption of the latter approach in this thesis included that a pure “clean-slate” MCDA for use in HTA could be associated with certain benefits as it can possibly address some theoretical and methodological limitations associated with the incorporation of costs and quality of evidence concerns. However, as demonstrated through one of the case studies (and also mentioned below), the application of MCDA approach that shares some characteristics to the supplementary or “incremental” mode might enable an easier exploration and implementation by decision-makers in real-world. Therefore, both approaches are associated with different advantages and disadvantages and as a result, the choice between the two approaches should be made depending on the decision context of interest, taking into account the current evaluation guidelines in place and the flexibility of the decision-makers.

Ultimately, the aggregate metric of value that is derived from the MCDA process is more encompassing because multiple dimensions of benefit are incorporated, together with their relative importance, as reflected by the preferences of the decision-makers. Adoption of MCDA-derived metrics as the benefit component can be used in combination with costs of the decision alternatives to estimate incremental cost value ratios (ICVRs) that could in turn inform resource allocation decisions on a similar basis to ICERs.

8.1.3 Methodological Construction of a New MCDA Value-Based Model and Framework

The development of a novel methodological framework for the evaluation of new medicines is described following an MCDA methodology based on MAVT, comprising a generic value-based model taking the form of a value tree (Advance Value Tree). A top-down “value-focused thinking” approach driven by the overall value concern (as part of an overall “value-alternative hybrid thinking” mode) was adopted for the construction of the value tree and in alignment with decision theory, which was informed from secondary and primary data collected from the literature and expert consultations as part of a five-stage process.

The value tree incorporated a number of criteria in a structured hierarchy form which can be adapted for different decision problems pertaining to different disease indications and therapeutic classes, aiming to capture various value concerns of decision-makers in the context of HTA that have traditionally been considered either explicitly in a systematic manner, or implicitly on an *ad hoc* basis. Value concerns were grouped into top-level criteria clusters relating to burden of disease, therapeutic impact, safety profile, innovation level and socioeconomic impact characteristics, which were then decomposed into mid-level criteria and bottom-level sub-criteria or attributes.

A precise combination of MCDA modelling techniques was proposed for operationalising the value tree through the construction of value judgements and elicitation of preferences as part of model assessment and model appraisal. An indirect elicitation technique based on pairwise qualitative comparisons and the use of value functions can be used for scoring the options (MACBETH), combined with an indirect swing weighting technique for eliciting relative criteria weights, and a simple additive technique for aggregating scores and weights together.

Altogether, the application of the selected MCDA techniques for the estimation of the generic value tree, while adopting a societal perspective which allows the incorporation of views from the wider stakeholder community, completed the development of the value framework (Advance Value Framework). Ultimately, it can enable decision-makers to understand their preferences and construct their perceptions for ranking and identifying the best decision alternative, therefore possibly acting as a decision support tool for decision-making.

8.1.4 Empirical Testing of the Value Framework with Multiple Stakeholders in England

The Advance Value Framework was tested in practice through a proof-of-concept case study adopting an integrated multi-criteria approach simulating an HTA context for the case of three metastatic colorectal cancer (mCRC) treatments. It engaged multiple stakeholders using the scope of the National Institute for Health and Care Excellence (NICE) in England, but in addition supplementary evidence was considered for value concerns not explicitly addressed by the Institute.

A bottom-up “alternative-focused thinking” approach driven by the decision alternatives was followed for adapting the Advance Value Tree and constructing a disease-specific value tree for the mCRC treatments in alignment with decision theory principles as part of the model-building phase. A facilitated decision analysis modelling approach was adopted for validating the value tree and eliciting the value preferences of thirteen participants through a decision conference, as part of the model assessment and model appraisal phases. Alternative treatments were indirectly scored across the criteria through the elicitation of value functions using MACBETH, an indirect MAVT elicitation technique. Relative weights were assigned to criteria using an indirect qualitative swing technique and overall weighted preference value (WPV) scores for the treatments were derived using a simple additive aggregation technique based on weighted averages.

This first application of the Advance Value Framework produced overall value rankings for three mCRC treatments across an explicit set of evaluation criteria through the incorporation of preferences from multiple stakeholders. Stakeholders’ experience suggested that the methodology adopted can aid the evaluation process by making it

more structured and transparent, however further research is recommended to enhance its use for policy-making.

One of the lessons learned was that the facilitated model-building approach used through decision conferencing and as part of the framework proved to be essential for creating a shared understanding of what constitutes value in the particular decision context. Under these circumstances of a heterogeneous participants group with multiple stakes, the consensus approach used for preference elicitation as part of an open discussion where debate was actively sought proved to be challenging at times but overall worth the effort. For such a consensus and shared understanding to be reached however appropriate facilitation turned out to be crucial, entailing the appropriate intervention to ensure that participants group remains on task and that weaker “voices” are also heard, therefore guiding the discussions but impartially without contributing to them. In addition, carrying out sensitivity analysis on the spot at the end of the workshop and exploring the impact of weight changes on overall options rankings proved to be helpful for resolving any remaining preference disagreements between the participants which might have persisted throughout the discussions, therefore helping to accept the final outcomes.

Another lesson was about the potential application of the methodology for the evaluation of medicines in other decision contexts. The value framework could be potentially beneficial if adapted accordingly and applied earlier or later in the life-cycle of new medicines. In the former case it could be used to enable the communication of the new product’s value with decision-makers when it might be still possible to amend some of its performance characteristics, allowing for a targeted redesign in the clinical development process (if feasible) and the collection of additional evidence if deemed necessary. In the latter case it could be used between clinicians and patients to enable the understanding of a medicine’s benefits and risks and construction of preferences around them, therefore acting as a catalysing tool in the shared decision-making process of treatment selection.

Overall, involvement of participants that were eager to share their views and challenge others’ as part of a shared commitment, turned out to be a crucial factor in the conduct of a successful multi-stakeholder evaluation using this value framework and with such a decision conference approach.

8.1.5 Empirical Testing of the Value Framework with Decision-Makers in Sweden

The Advance Value Framework was tested in practice through another proof-of-concept case study adopting a theory-based integrated multi-criteria approach for assessing the value of three metastatic castration resistant prostate cancer (mCRPC) treatments. It engaged real world decision-makers using the scope of the Swedish Dental and Pharmaceutical Benefits Agency (TLV) in Sweden, but in addition supplementary evidence was considered for value concerns not explicitly addressed by the Agency.

A bottom-up “alternative-focused thinking” approach driven by the decision alternatives was followed for adapting the Advance Value Tree and constructing a disease-specific value tree for the mCRPC treatments in alignment with decision theory principles as part of the model-building phase, feeding from the literature and expert opinion. A facilitated decision analysis modelling approach was adopted for validating the value tree and eliciting the value preferences of four participants through a decision conference, as part of the model assessment and model appraisal phases. Alternative treatments were indirectly scored across the criteria through the elicitation of value functions using MACBETH, an indirect MAVT elicitation technique. Relative weights were assigned to criteria using an indirect qualitative swing technique, and overall weighted preference value (WPV) scores for the treatments were derived using a simple additive aggregation technique based on weighted averages.

This second application of the Advance Value Framework produced overall value rankings for three mCRPC treatments across an explicit set of evaluation criteria through the incorporation of preferences from decision-makers. Decision-makers’ experience suggested that the methodology adopted has the prospects of facilitating their decision-making because it allows for explicitly assessing performance across a multiplicity of criteria, while incorporating trade-offs on their importance, in a structured and transparent way. However further research is recommended to address technical difficulties and advance its use for informing policy-making.

Having gained clarity on the preference elicitation and the underlying construction process, the final results were accepted from the group as a whole due to the transparency and the step-wise manner characterising the methodological process. The systematic elicitation of preference across all attributes individually, followed by the visual presentation of the options’ respective performance, enabled the participants to develop a holistic and confident understanding around the overall strengths and

weaknesses of the options. As a result, distinguishing between the performance of the options and the valuation of these performances, as operationalised through the indirect scoring and weighting techniques used, was critical.

Among the lessons learned was that for the possible implementation of the methodology to be catalysed even at a testing phase, the development of the model, and the application of the value framework altogether, should be aligned with the formal remit of the decision-makers. Establishing clear goals for the role of the results would be beneficial for the initial exploration of the methodology. In settings where strict guidelines and requirements exist around the application of specific HTA methods as in Sweden with the use of economic evaluation, the exploration of a supplementary “incremental” MCDA mode to CEA might be more attractive and realistic option to decision-makers so that the results can be more readily adaptable to their current needs and therefore being more easy to implement in their daily practice. In turn, adaptation of a pure “clean-slate” mode could still be viable but involving the assessment of more concise value trees that restrict the number of criteria around the formal value concerns currently in explicit use. This would allow capturing the primary value concerns as specified by the evaluation guidelines and requirements while still offering the added benefits of analysing value through a decomposition approach that could enable the facilitation of decision-making through the step-wise and transparent elicitation and construction of value preferences. Following such a testing phase, decision-makers could acquire important confidence that could enable the exploration of incorporating additional value concerns, therefore catalysing a shift to change any institutional remit in place.

8.1.6 Overall Thesis Conclusion and Contributions

Assessing and appraising the value of new and expensive medicines and health care interventions in general acts as a major challenge for health care systems especially in settings of significant budgetary pressure. Novel methodological approaches for the evaluation of new medicines are urgently needed that can improve the HTA decision-making process to support better decisions, to improve efficiency in resource allocation and therefore drive an increase in population health.

This thesis proposes that evaluation procedures aiming to rank alternative treatment options should be characterised by extensive comprehensiveness and full

transparency, enabling the structured elicitation and explicit incorporation of preferences across multiple value dimensions and from a range of stakeholders, for the decision-making process to be rational and their outcomes credibly accepted.

The key research question in this thesis is whether - and how - MCDA could provide the basis of an alternative methodological framework for assessing the overall value of new medical technologies by explicitly capturing all relevant value aspects. The overall aim of the thesis is decomposed into a number of secondary research questions relating to: i) the design of a robust MCDA methodological process for HTA, ii) the identification of value parameters that are applied as evaluation criteria in HTA across different EU countries, iii) the development of a value-based model that can capture multiple value dimensions for the evaluation of new medicines in HTA as part of a broader MCDA methodological framework and in alignment with decision theory, and iv) the adaptive application of this value framework by means of two case studies to assess and rank a set of alternative treatment options allowing the elicitation of stakeholder and decision-makers preferences that could inform coverage decisions.

The secondary research questions are addressed separately through the five empirical papers of the thesis. Initially a structured MCDA methodological process was constructed based on MAVT for the context of HTA. Then, a systematic literature review and expert consultation was conducted to study current HTA practices in Europe, identifying value concerns of decision-makers and evaluation criteria in HTA while highlighting the limitations of the methodological approaches in place. These results fed into the development of a MAVT methodological framework, incorporating a value-based model taking the form of a generic value tree that can be adapted to capture decision-makers' value concerns, operationalised through a combination of MCDA modelling techniques. Finally, the new value framework was applied and tested in practice through two case studies involving the participation of stakeholders and decision-makers.

As a result, the contribution of the thesis comprises conceptual, theoretical, methodological and empirical components. The completion of the tasks in the different thesis papers required me to transcend the current HTA methods I had been exposed to, making it essential to acquire, develop and apply important interdisciplinary skills. An extensive amount of work had to be completed, involving a wide range of tasks relating to understanding and gaining knowledge, conceptual and critical thinking, theoretical construction, review of the literature, conduct of semi-structured interviews,

consultation with experts, organisation and facilitation of decision conferences, analysing results, and policy recommendations.

Ultimately, the methodological framework produced could enable HTA decision-makers to understand and construct their value perceptions and preferences for the purpose of assessing, ranking and identifying the best decision alternative, therefore acting as a decision support tool for facilitating their decision-making process. Future research could inform the validation of the value framework's robustness while possibly showcasing its usefulness through other practical applications involving HTA bodies and decision-makers across different countries while engaging multiple stakeholders. Importantly, the application of the value framework and its methodology could be explored in other decision contexts along the life-cycle of new medicines and health care interventions. Application for early pre-marketing evaluation as part of clinical development could enable the communication of value to decision-makers and help streamline collection and preparation of evidence requirement, whereas use by clinicians and patients could assist them with understanding the benefits and risks of new treatments and aid their selection process.

8.2 Policy Implications

The MCDA methodological process (Chapter 3 – Paper 1) resulted in the development of a value framework for the assessment of new medicines in the context of HTA (Chapters 4 and 5 – Papers 2 and 3). The value framework was tested in practice with two real-world applications in the context of coverage decisions: a case study on metastatic colorectal cancer with multiple stakeholders while adopting the perspective of the English HTA agency (NICE) for assessing and ranking the overall value of second line biological treatments, and a case study on metastatic prostate cancer with the Swedish HTA agency (TLV) to assess and rank the overall value of a set of treatments from the viewpoint of the agency.

The case studies proved to be successful applications of the MCDA methodological framework in practice for the context of HTA from the perspective of the stakeholders and decision-makers. Overall, the participants of the workshops felt these were very interesting simulation exercises with useful insights and with the actual process proving to be easier and less complicated than originally expected.

Perceived benefits of the value framework included the explicit incorporation of multiple “non-traditional” value dimensions, especially innovation related attributes,

the facilitation in recognising and expressing the relevance of different value concerns for the particular decision context and the ability to illustrate differences in views within the group. In addition, the overall process enabled participants to realise that it is possible to consider multiple value dimensions within the HTA process (for which they were positive about) and that an explicit and quantitative performance measurement turned out to be feasible for a range of value concerns for which originally only implicit and qualitative assessment approaches were thought to be applicable. Furthermore, besides the prospects of acting as a decision-making tool it was suggested that the value framework could also be used as an important negotiation tool during the discussions with the manufacturers and as part of the early dialogue process, aiding communication and “signalling” their preferences down to them due to its comprehensiveness and transparency.

8.2.1 Reflections on the Link between Value and Policy-Making

Assuming the proposed methodological process is adhered to, the application of MCDA presents a number of advantages to decision-makers in the context of HTA and the wider context of value based assessment compared with currently used HTA approaches such as economic evaluation techniques.

Firstly, it acts as an instrument of more complete value assessment leading to improved comprehensiveness given the explicit incorporation of multiple criteria and construction of value judgements on the performance of alternative options that can help decision-makers to construct their overall value perceptions. Secondly, assignment of quantitative criteria weights can reflect differences in the relative importance of the evaluation criteria, enabling decision-makers to realise the value trade-offs they are willing to make and therefore understand their value preferences. Thirdly, the methodological process can be informed through extensive expert engagement and direct stakeholder participation leading to an encompassing nature of value perceptions and value preferences. Fourthly, it provides flexibility given that the details and technical characteristics of the different methodological stages can be adapted to accommodate particular decision-makers’ needs. Finally, the entire process is fully transparent, allowing to illustrate the rationale behind the decision outcomes which could help them become credible and accepted from the wider stakeholder community and society.

Therefore an important difference between economic evaluation methods such as CEA and MCDA is that the latter facilitates a decision support system, as illustrated in Figure 8.1. As can be viewed from the top of the Figure, in CEA the analysis of costs and health gains is first taking place explicitly as part of the assessment process (left-hand side), and value trade-offs are then elicited implicitly possibly involving other types of benefit gains and stakeholder views on an *ad hoc* basis as part of the appraisal process (right-hand side), so that a decision can ultimately be made. In contrast, as evident from the bottom of the Figure, in MCDA the analysis for all types of benefit gains and their value trade-offs as informed through stakeholder views are altogether explicitly incorporated in the overall process, thus better linking assessment and appraisal of evidence and supporting decision-making.

The resulting overall WPV scores derived from the MCDA process can act as a more encompassing measure of value given that multiple benefit dimensions are explicitly assessed and therefore could be used to drive the coverage decision and pricing negotiations of new medicines and health care interventions in a more comprehensive manner. Consideration of purchasing costs in parallel with the overall value of the alternatives options can then be used to estimate incremental cost value ratios (ICVRs) for the different health care interventions and contribute to priority setting and resource allocation decisions in the context of HTA. For example, funding for the coverage of a set of interventions could be allocated based on their ICVRs rankings, from the lowest ratio to the highest ratio, until the available budget is exhausted. In turn, budget impact considerations could be incorporated in the cost-value ratio estimates at system level by taking into account the expected number of patients receiving the respective interventions.

In this context, the case studies conducted aimed to assess and rank alternative treatment options for the same disease indication. A number of disease-specific clinical endpoints were incorporated as evaluation criteria as they reflected a number of common value concerns relating to the particular intra-indication decision context of interest. Such an attempt at a broader inter-indication level, aiming to assess the value of alternative treatments for the purpose of different disease indications might be more challenging as it would have to ensure the use of a common value model (in terms of attributes, value functions and relative weights) that adequately addresses the value concerns for the alternative treatments across diseases with different characteristics. In this instance, criteria and attributes might need to become more generic and less

disease-specific, using health benefit metrics such as the QALY, in which case trade-offs might have to take place between the sensitivity/specificity of assessing the different treatments and the comprehensiveness of the analysis.

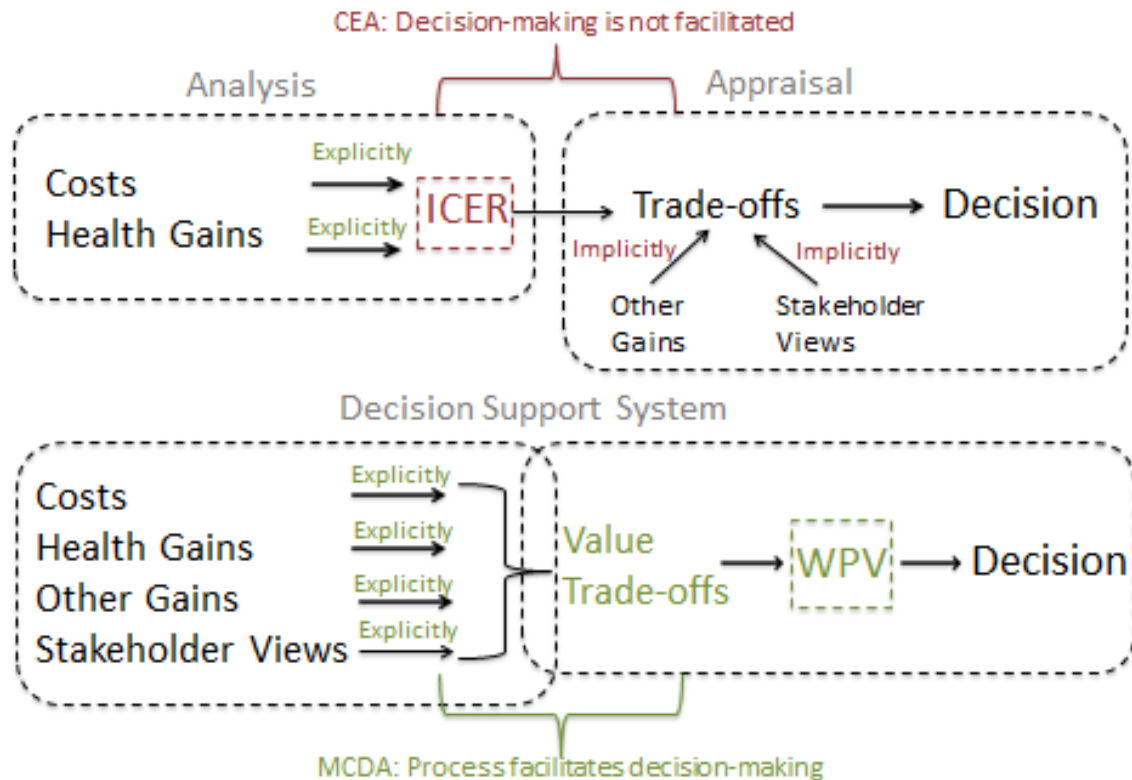
Ultimately, because of its characteristics enabling a structured process, MCDA could facilitate overall decision-making acting as a decision-support system, enabling its use as a reasonable resource allocation tool that, among others, incorporates a more holistic approach to value.

8.2.2 Responding to HTA Issues at Systems Level: From Methodological Robustness to Practical Relevance

So far, decision-makers and even health care evaluation researchers exploring the application of MCDA-like approaches might not pay adequate attention to the theoretical foundations and good practices of MCDA as for example the different set of properties that the multi criteria evaluation models need to possess for the analysis to be robust. Recent evidence has shown that only one healthcare MCDA study explained that criteria were defined to meet MCDA requirements such as avoiding double counting ⁴⁸⁷, with others acknowledging as a concern the fact that MCDA responders might not understand some of the attributes being used ⁸⁴, possibly because of difficulties in interpreting the meaning of the respective attribute performance ⁴⁸⁸.

Taking into consideration that MCDA in itself posits a departure from currently used HTA techniques, the application of an MCDA approach and its principles in the context of HTA requires careful reflection on a number of fronts.

Figure 8.1: Differences between cost effectiveness analysis (CEA) and multiple criteria decision analysis (MCDA) in decision-making



Firstly, it is important to clarify whose preferences to consider. Assuming that an HTA agency acts as a proxy decision-maker, then it would be appropriate to adopt the perspective of the respective HTA agency. For example, if the decision context is England, France or Sweden, it would be reasonable to adopt the perspective of the National Institute for Health and Care Excellence (NICE), the Haute Autorite de Sante (HAS) and the Dental and Pharmaceutical Benefits Board (TLV), respectively. Consequently, any social judgements individual HTA agencies adopt, including the participants and their preferences, will need to reflect the particularities of each setting. As different countries or settings are likely to have different priorities and objectives, the analysis should be tailormade to their needs. Alternatively, if the adoption of such an existing perspective is not possible, some formal stakeholder analysis could be used to identify the key players that should be involved ⁴⁸⁹.

A second, but related issue, is how to combine the preferences of individual stakeholders. Ideally, a consensus approach should be aimed for, through which a single agreed value judgement (i.e. score, weight) would be derived. Alternatively, if mathematical aggregation is used, the median (or mean) of the responders' preferences

could be used, especially in settings where motivational biases from some stakeholders exist (i.e. strongly against vs. strongly in favour). In any case, the complete range of value judgements should be recorded and be used for the purpose of sensitivity analysis, where the impact of different scores and weights on the options' total value scores would be tested.

A third practical issue relates to the evidence requirements and its availability. While MCDA has been criticised that it requires more evidence than standard HTA approaches in order to populate the criteria, in practice the same evidence that is required for standard HTA approaches can be used in the context of an MCDA model. Even if certain items of information are not available, they should be readily collectable, or at least able to be proxied through expert opinion.

The consistency of results would be a fourth important issue subject to criticism. Would the results be consistent within the same setting or could inconsistencies act as an obstacle to homogeneous decisions? If the value analysis is evidence-based, the results are likely not to be identical if the participants whose preferences are considered have different value judgements (i.e. different value functions and weights), but it would be expected to be similar. This highlights the importance of the context where MCDA methods should be applied. Indeed, they should act mainly as decision aiding tools to support health care resource priority setting conducted by the decision-maker.

8.2.3 Roadmap to HTA Application and Implementation

Given that MCDA is a departure from conventional HTA approaches, a roadmap would be needed on how MCDA could be factored in current HTA practices (Figure 8.2). To begin with, any efforts of MCDA implementation should start by building a research team with the appropriate technical expertise as part of an education phase. A multi-disciplinary research team spanning the fields of decision analysis, medical and life sciences, health economics, and statistics would be recommended.

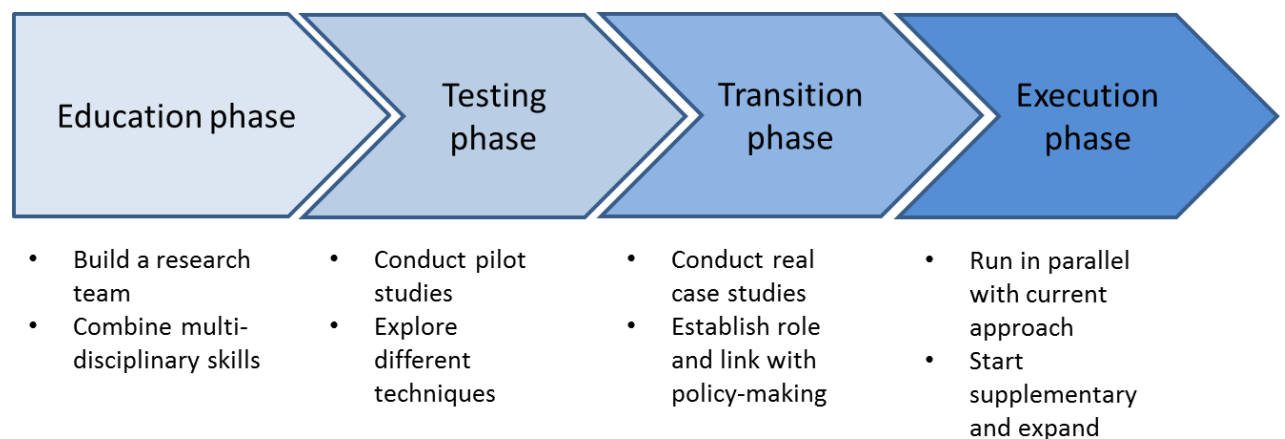
A number of pilot studies could be carried out in a testing phase. These pilots would act as testing exercises aiming to simulate the evaluation process and the production of hypothetical HTA decisions, in order for the Agency personnel to gain a first-hand experience on the technical aspects of the MCDA process. A variety of MCDA methods and techniques could be explored, following which a range of

techniques would be short-listed for future use based on their relative advantages and disadvantages.

Actual case studies could then be conducted as part of a transition phase, using real evidence from past health technology appraisals. Value judgements and preferences could be elicited from actual members of the Agency's appraisal committees while acting as participants in decision conferences taking the form of facilitated workshops. The results could be used to highlight any differences in recommendations with past appraisals, helping to realise the benefits and insight of the methodology as part of real practice. This phase could end with the establishment of clear role(s) for the envisaged use of the new methodology and its link with policy-making.

Finally, the MCDA approach could become fully operational as part of an execution phase, running in parallel with any existing formal appraisals taking place. The MCDA approach could start as a supplementary source of information, acting as a decision-making tool on top of standard appraisals, and then following the decision-making needs and vision of the Agency it could eventually become the sole approach implemented.

Figure 8.2: Proposed roadmap to MCDA application and implementation



8.3 Limitations

8.3.1 Conceptual and Theoretical Limitations

Although the methodology proposed in this thesis is based on value measurement methods and precise multi-attribute value theory (MAVT) methods, other methods

from different “schools of thought” are also available such as satisficing and aspiration levels methods, outranking methods and fuzzy and rough sets methods. The MAVT methodology was chosen mainly because of its practical simplicity, compensatory nature, relevance to HTA, and popularity with health care applications altogether. In any case, the systematic and more extensive exploration of the other MCDA methodologies could help to more accurately identify different advantages and disadvantages between these methods, which could make the adoption of alternative methodologies preferred for the case of particular decision contexts or decision-makers’ needs and peculiarities.

Consequently, indirect preference elicitation techniques were adopted for the purpose of scoring and weighting, mainly because of their theoretical robustness and unbiased nature. More precisely Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) was adopted because of its strong theoretical foundations and convenience for decision-makers, as part of which qualitative judgements about difference of value are constructed to elicit value functions and assign relative weights through indirect qualitative swing weighting. However other MCDA modelling techniques of varying complexity exist for the formation of value judgements and elicitation of value preferences, some of which might have been preferred by others. Subsequently, in terms of the generic value tree model developed and used as part of the value framework, I tried to identify and capture decision-makers concerns as comprehensively and objectively as possible, through a five-stage process involving literature reviews and consultations with experts. I focused on the value concerns of European decision-makers, and although the model was consulted with experts from Latin America and Eastern Europe, it could still turn out not being comprehensive enough at global level for some regions or contexts. However, it should be relatively straight-forward to tailor it to any missed decision-maker needs through some adaptation or restructuring.

Last but not least, another challenge would relate to the next possible step required for MCDA implementation in policy making: setting up a hypothetical efficiency cut-off point, assuming it is needed, essentially an alternative to the current incremental cost effectiveness ratio (ICER) threshold, or in other words an incremental cost value ratio (ICVR) threshold. Such an attempt would not be limited to the application of MCDA and would face all the theoretical and practical hurdles associated with the estimation of a sound ICER threshold based on opportunity cost that have been

seen to date ^{485,486}. In addition, it might also have to ensure that a common value model is used, inclusive of attributes and their value functions as well as criteria weights. Although this could be relatively easy for intra-indication (i.e. same disease) evaluations, an example of which would be the current case study, this would be more challenging for inter-indication (i.e. different disease) evaluations. In this latter case, the respective attributes might need to be less disease-specific and of a more generic nature, therefore leading to potential trade-offs between the sensitivity/specificity of assessing the different treatments and the comprehensiveness of the analysis.

8.3.2 Empirical and Methodological Limitations Based on the Case Studies

Results from the application of the value framework through the case studies should be interpreted with caution. It should be clear that these are simulation exercises aiming to test the new framework in practice and to illustrate its application and not to inform policy making in this instance in respect to the particular decision problem under consideration.

Given the absence of head to head clinical trials comparing directly all treatments of interest and the existence of relative treatment effects across the clinical attributes of interest, un-synthesized evidence from the respective single pivotal trials of the alternative treatments were used. However, using evidence from different clinical trials to directly compare alternative treatments is not accurate, even if the populations of patients across the different trials are similar (in terms of disease severity and treatment history). Ideally, an indirect treatment comparison should be conducted first using a common comparator to estimate the relative effects of two treatments versus the comparator, or even a network meta-analysis that combines both direct and indirect evidence available through a mixed treatment comparison ⁴⁹⁰. Such a common comparator was absent among the pivotal trials under consideration and conducting a meta-analysis was outside the scope of the study.

Therefore, an important limitation of this study is the lack of relative effects as part of clinical evidence and the use of their absolute effects from different clinical trials with the assumption that they can be directly compared. In real world evaluations aiming to inform policy-making decisions, evidence synthesis would be required to take place together with evidence collection as part of the model-building phase. An example of such an evidence synthesis stage would be the application of an SMAA

approach for assessing the comparative benefit-risk of alternative statins in primary prevention³⁴⁰, using comparative effects from evidence of three meta-analyses⁴⁹¹⁻⁴⁹³, or the combination of SMAA with a network meta-analysis for assessing the comparative benefit-risk of second-generation antidepressants⁴⁹⁴.

An important challenge would relate to technical difficulties associated with ensuring that all attributes possess the needed properties for a multi-criteria evaluation, and particularly that they are preference-independent. Preference independence between the values of the attributes is a necessary property for the use of simple additive models like the one used. Also referred to as “difference independence”⁴⁹⁵, this notion denotes that the difference in attractiveness – or added value of an improvement – between different attribute levels does not depend on the measurement of other attributes⁴⁹⁶. Such a “difference consistency” between (the values of the) attributes was taken as a working hypothesis given that it is “so intuitively appealing that it could simply be assumed to hold in most practical applications”⁴⁹⁷ (p. 284).

However, in the course of the decision conferences it became evident that the criteria attributes of OS and HRQoL might not be preference independent. More precisely, during the elicitation of the OS value function but also during the swing weighting stage, it became apparent that in order to indicate the magnitude of value associated with an extension in OS one might need to consider the respective HRQoL accompanying it; however the reverse was not evident, when the HRQoL value function was elicited. This issue was encountered by revealing the “lower” (x_l) and “higher” (x_h) reference levels of the dependent HRQoL attribute (EQ-5D index scores), i.e. 0.75 to 0.86, and 0.72 to 0.82 for the mCRC and mCRPC case studies respectively, and instructing the participants to assume an identical performance between the options on the dependent attribute (given that all three options had identical EQ-5D index scores), but without disclosing the exact figure. In other words, during the elicitation of the OS value functions the “lower-higher” range of the *HRQoL* attribute was revealed, i.e. $v(x_{0.72}) = 0$ and $v(x_{0.82}) = 100$ for the case of mCRC case study, and $v(x_{0.72}) = 0$ and $v(x_{0.82}) = 100$ for the case of the mCRPC case study, while acknowledging that the performances of the options were indifferent (i.e. identical), therefore deriving an OS value function that was effectively conditional on the range of the HRQoL attribute.

In order to be 100% theoretically correct, existence of such preference dependence would normally require the use of more complex multilinear (i.e. multiplicative) models, given that the synergy or antagonism between the two

respective attributes would have to be reflected in the model accordingly. In order to avoid using such a less straightforward model and keep using a simple additive one, the synergy (or antagonism) effect between the two attributes would have to equal zero, which would hold true if the product of their value scores is zero. Therefore, in a secondary analysis the value scores of the options across the *HRQOL* attribute were changed to zero (given their indifference in their performance), which did not result in influencing their ranking. Another alternative to avoid using a more complicated model would be to just eliminate the *HRQOL* attribute from the model given the indifference in the options' performance. As part of another secondary analysis, the *HRQOL* attribute was removed from the model, without impacting the ranking of the options either. It should be noted though that any of the above modifications would only be possible because of the indifference arising in *HRQOL*, and in the case that the *HRQoL* performances of the options were not identical then it would be required that the *OS* and *HRQoL* attributes get combined into a single aggregated attribute for the simple additive model to be used. A recommendation for future MCDA applications evaluating treatments for mCRC, mCRPC and oncology indications in general or even end-of life situations would therefore be to explore potential preference (i.e. difference) dependence between the measures of *OS* and *HRQOL*, validating any difference (in) consistency, while possibly also investigating the complete attribute ranges for which it holds true, therefore giving an insight on any conditions in place.

Furthermore, in the present case studies only the HRQoL of the stable disease state was assessed because none of the treatments were assumed to have any effects during the progressive disease state^{427,442,467}. However in other disease indications this might not hold true in which case the relevant HRQoL attribute would need to capture both the stable and progressive disease states, possibly in the form of an OS-HRQoL aggregated attribute (given their possible preference dependency), therefore producing a similar metric to QALYs.

Also, among the main challenges of the methodological approach adopted is the relatively subjective nature of setting the “higher and lower” reference levels on each attribute, based on which treatment scores are derived. For example, as explained above, in the case of the treatment discontinuation attribute of the mCRPC case study, this was not set equal to 20% less than the best performing option (the reverse logic to the case of the therapeutic impact attributes), but rather equal to the minimum natural

limit of the scale (i.e. 0%) which was regarded as an “ideal” level. This level could indeed be perceived as “too extreme”, or “too good to be true”.

Possibly more important however in terms of the treatment discontinuation attribute’s impact on the scoring of the alternative treatments was the definition of the “lower” reference level as this could influence the negative performance scores observed in two of the treatments and consequently their overall WPV scores. The lower reference level of 10% adopted on the basis of Best Supportive Care (i.e. “doing nothing”) performance was sourced from the placebo comparator arm of enzalutamide’s pivotal clinical trial (*AFFIRM*). The choice of using the placebo arm of the *AFFIRM* trial to proxy BSC performance and not the comparator arm from any of the other two treatment’s pivotal trials that were used in the analysis was because it better resembled BSC; in abiraterone’s pivotal trial (*COU-AA-301*) all patients in the placebo comparator arm were also administered steroids (prednisone), whereas in cabazitaxel’s pivotal trial (*TROPIC*) all patients in the comparator arm were administered steroids (prednisone or prednisolone) on top of chemotherapy (mitoxantrone). As a result, abiraterone and cabazitaxel produced negative preference value scores in treatment discontinuation as their performance lied below the lower reference level.

Setting reference levels that were not necessarily equal to the limits of the value scale, ensured that the scale had enough granularity to distinguish the treatments, which is not always the case if the maximum and minimum levels are employed as reference levels. The basis adopted for setting these levels are clearly and extensively described in the respective methods and results sections (Chapters 7 and 8), and although I tried to be as objective as possible, others would most probably have ended up with different points as anchors. However, such differences would most probably be minor, not necessarily affecting the overall valuation of the treatments. The final reference points adopted were decided following the feedback that was received during the workshop for the case of the mCRC case study and from a clinical specialist for the case of mCRPC case study. Liaising with a range of experts during the model building phase for ensuring the choice of relevant reference levels, possibly before the actual model assessment phase, seems to be a necessary step for ensuring good practice and robust results.

Another challenge would be the evaluation of clusters where participants have less experience or knowledge. For example, during the evaluation of the *HRQoL*

attribute in the mCRC case study, some of the participants had difficulties in comprehending the differences in value between the different QALY (i.e. EQ-5D index) scores. However, feedback from clinicians and patients helped the rest of the group to understand the relative differences across health states so that they could express their preferences. Although such an input proved crucial and to a large extent satisfactory to the information needs of the group, it would be advantageous if the QALY scores were accompanied with descriptions across the EQ-5D dimensions for the given scores.

A couple of other issues were perceived by some of the workshop participants to act as potential limitations in the mCRPC case study. One opinion was that too much attention was paid to details of the model that eventually exerted no major influence on the overall WPV scores or any actual impact on the options' rankings therefore possibly being redundant, as for example with the case of Special Instructions and Delivery Posology attributes. However, although the relative importance of some attributes might ultimately turn out to be relatively small, this cannot be predicted before preferences elicitation and even in this case their combined weight might still add up to be critical; in the example of the Special Instructions and Delivery Posology attributes, the sum of their relative weights was 7.3% of the total which should not be regarded as insignificant and in other instances it could have affected the options' rankings.

The view was also expressed that cost per QALY is still favoured by health economists and therefore whether it might be preferable to use “number of QALYs gained” as the basis of value, on top of which other additional benefit components could then be added. The main logic behind decision analysis and MCDA is to decompose complicated decision problems into simpler problem components, so that they can be analysed separately and then combined to inform an action plan for the problem, essentially adopting a “divide and conquer” approach ⁵³. In the HTA context, where decision problems relate to the coverage of new drugs, such an approach would therefore require the evaluation of the alternative interventions against all relevant value dimensions individually (if possible). As a result, considering and analysing the performance of the alternative interventions against OS and HRQoL distinctively should aim in helping decision-makers express their value judgements and understand their overall value preferences across the options. However as mentioned above, a plausible preference dependency observed between the OS and HRQoL attributes could

indeed require their combination into a single aggregated attribute of benefit resembling number of QALYs gained.

8.3.3 Cognitive and Motivational Biases

Importantly, research on behavioural decision analysis has indicated that construction of judgements as part of decision-making is prone to a number of biases, relating to faulty cognitive processes or due to motivations for preferred analysis outcomes ⁴⁹⁸. The presence of these biases could be present across the different phases of the methodological process developed and adopted, and especially in the model-building, model-assessment, and model-appraisal phases, that took place as part of the case studies involving the participation of experts and decision-makers. Such biases could potentially reduce the quality of the model and the results of the analysis.

A number of de-biasing techniques have been recommended to overcome these limitations⁴⁹⁸ and although many of them were applied during the decision conference workshops as part of the specific modelling and preference elicitation tasks, important sources of biases could still remain. For example, in terms of cognitive biases, ‘equalizing bias’ relating to the allocation of similar weights to all value concerns was addressed through the hierarchic elicitation of weights. ‘Gain-loss bias’ which occurs as alternative descriptions of a choice and its outcomes (which may lead to different answers) was addressed by conducting value judgements in relevance to (marginal change from a) best supportive care option which was used as a reference levels and which could act as a *status quo*. ‘Myopic problem representation bias’ which occurs when an oversimplified problem is adopted and based on an incomplete mental model for the decision problem was addressed by explicitly encouraging experts to think about more value concerns in the wider socioeconomic context. ‘Omission of important variables bias’ was addressed by using group elicitation techniques so that no important variable is overlooked. ‘Overconfidence bias’ relating to overestimation and overprecision was addressed by using fixed value instead of fixed probability elicitations. ‘Proxy bias’ which occurs when proxy attributes receive larger weights than the respective fundamental concerns was addressed by trying to avoid the use of proxy attributes. ‘Range insensitivity bias’ occurring when objectives are not properly adjusted to changes in the range of attributes was addressed by making attribute ranges explicit and using a swing weighting technique. ‘Splitting bias’ which occurs when the

structuring of the criteria affects their weights was addressed using hierarchal estimation of relative weights.

Then in terms of motivational biases, a number of biases could exist because of peoples' emotions, desires and motives ⁴⁹⁸: 'affect influenced bias' relating to the emotional predisposition for or against a specific outcome, 'confirmation bias' relating to the desire to confirm one's belief leading to unconscious selectivity, 'desirability of a positive event or consequence bias' occurring when the desirability of an outcome leads to an increase in the extent to which it is expected to take place (i.e. 'wishful thinking' or 'optimism bias'), 'under-desirability of a negative event or consequence bias' occurring when there is a desire to be cautious or conservative in estimates that may be related to harmful consequences, and 'desirability of options or choice bias' leading to over- or underestimating values or weights in a direction that favours a desired alternative. All of these were addressed by engaging multiple experts with alternative points of view, collecting views from a range of different experts as stakeholders or decision-makers to provide different value perspectives, and using indirect MAVT techniques for scoring the options and weighting the criteria. Therefore, the use of a precise combination of indirect MAVT methods and more precisely swing weighting hierarchic elicitation techniques at group level, taking place through participatory processes involving key stakeholders and experts which are facilitated appropriately such as through Decision Conferencing, could act as an important way for avoiding the occurrence of many detrimental cognitive and motivations biases in the context of HTA which could deteriorate results' robustness.

8.4 Future Research Directions

Future research could explore three distinct, but interconnected areas. The first would be to conduct further empirical applications of the value framework in the context of HTA with the view to validating its use for the purpose of reimbursement and pricing decisions and as a resource allocation tool and, ultimately adapting it to the needs of decision-makers. Second, to expand the application of the framework in the decision context preceding HTA (notably for early HTA) and also for licencing approvals, but even as part of drug development during clinical trials. And, third, to adapt the framework in the context of treatment selection and prescribing between patients and clinicians, known as shared decision-making.

8.4.1 Validation in the HTA Context

In terms of generating further empirical evidence in the context of HTA applications, a number of case studies could be conducted in different countries with a number of different public HTA agencies across Europe. Such a research initiative could seek to tailor the framework as a methodological tool that can capture the preferences of the decision-makers based on their needs in a robust and transparent manner. In doing so, the results from a number of identical evaluation contexts across countries using the same evidence could be compared in order to highlight the differences in the value judgements and preferences of different European payers/regulators, therefore giving an insight of important variations in the regulatory landscape in a cross-country comparison fashion.

For example, together with my PhD supervisor Prof Panos Kanavos, I am currently in the process of organising similar case studies with different HTA agencies and payers across Europe with intention of advancing the framework as a tool for the needs of decision-makers. I have already conducted workshops with the Andalusian Health Technology Assessment Agency (AETSA), the Polish Agency for Health Technology Assessment and Tariff System (AOTMiT), and the Belgian National Institute for Health and Disability Insurance (INAMI-RIZIV) and I am looking forward for others to follow.

Other countries could include England, France, Italy and Netherlands with the involvement of the national HTA agencies/insurance funds of each country (i.e. NICE, HAS, AIFA and NIZ respectively). An additional research objective would be to test the methodology with private health insurers based in the US. Given the high cost and uncertainty of cancer treatments, in tandem with the high burden of disease, a set of oncology indications such as prostate, lung, breast and skin cancer could be included for which a number of biological molecules and immunotherapies would be evaluated. Selected indications could also include less prevalent cancers with higher unmet medical needs such as rare blood cancers (e.g. chronic myeloid leukaemia, myelofibrosis), in order to eventually compare and contrast the evaluation process of orphan medicines versus non-orphan medicines. Assuming a relatively high number of case studies across different countries will be conducted, the methodology could be adapted for the preferences to be elicited in a virtual mode, requiring the development of a web-based application. Importantly, the existence of different types of behavioural

biases in these specific contexts as described in section 8.3.3 could be explored, together with possible techniques for addressing them.

8.4.2 Application into Drug Development and Early HTA Context

In terms of expanding its application earlier in drug development, the methodological framework could be applied in the benefit-risk assessment taking place for the purpose of licencing approval as part of the traditional marketing authorisation process, but also as part of more innovative adaptive approaches to marketing authorisation which depend on continuous evidence development. Besides that, the framework could be adapted for evidence generated in phase 2 or 3 clinical trials to evaluate molecules earlier in the product-lifecycle in order to better communicate future value prospects with decision-makers - both regulators and payers - while improving alignment with evidence requirements⁴⁹⁹. In this context, the framework could be used as a tool for facilitating the expression of value concerns and the construction of value preferences of decision-makers, with the analysis including HTA (and licencing) related criteria that reflect payer (and regulator) concerns, but earlier on, when the evidence generation process as part of drug development is still amenable. This could help address the gap between what payers and HTA bodies perceive as value and what the industry delivers following marketing authorisation requirements, helping to ensure that the envisaged clinical trial evidence generated for the purpose of licensing approval meets payer and HTA body requirements. Eventually this could help to optimise the overall drug development process and bring down expenditures by identifying redundant studies/outcomes that are necessary and might be currently missing. Given the high probability of failure and the high unmet need, possible case studies could use evidence from CNS products relating to neurological conditions such as Alzheimer's, Multiple Sclerosis and Parkinson's.

Such a research would be in alignment with current initiatives from the European Medicines Agency (EMA), relating to pilot projects that explore the use of adaptive pathways and the cooperation between regulators and payers/HTA bodies.

In terms of adaptive pathways, this is a product development concept for medicines targeting unmet needs which could be defined as “a planned, progressive approach to bringing a medicine to patients”⁵⁰⁰. Under such a pathway, a new medicine would first receive an authorisation for a relatively small population of patients that is

likely to benefit the most and over time progressive licencing approvals could follow, extending the original indication through the collection of additional evidence in wider patient populations. The first such pilot project was just completed and a number of aspects were identified for further reflection, including the need for increased patient involvement, the potential to involve payers in order to provide input on pricing strategies and the definition of methodologically sound strategies of real-world evidence collection. All of these issues could potentially benefit from the application of MCDA approaches: involvement of patients and payers could inform the incorporation of patient and payer related concerns in the evaluation process and the elicitation of their value preferences, and flexible models could be constructed that are easily adaptable for the accommodation of future evidence collection.

In terms of cooperation between regulators and HTA bodies, EMA together with EUnetHTA recently completed collaboration as part of a work plan with various aims and achievements. These included the establishment of early dialogue for medicines manufacturers to reduce duplication while streamlining and enhancing the whole development process, improving EMA assessment reports to address the requirements of HTA bodies, and the facilitation of rapid relative effectiveness assessment⁵⁰¹, which in turn could be addressed via the application of the framework for the elicitation and construction of value preferences aiding value communication and evidence alignment as mentioned above.

8.4.3 Application in Shared Decision-Making

With regards to the treatment prescribing application, the framework could be adapted in order to aid the selection process of the most appropriate treatment at patient level as part of a shared decision-making context involving patients and clinicians⁵⁰². Patient centeredness and engagement has been emerging as a vital way for achieving better health outcomes and efficient allocation of resources, being especially relevant for settings with high availability of treatment options and with differences in their benefits and risks, usually requiring elicitation of benefit-risk trade-offs. This application could therefore explore clinical prescribing contexts in which multiple treatments for the same disease indication are routinely prescribed in clinical practice but which might produce heterogeneous outcomes in different patient applications. Heart disease could be selected as the disease indication for which a number of different

statins would be assessed against based on their comparative benefit-risk ratios across a range of health outcomes⁵⁰³. The methodology could be adapted so that preferences can be elicited from a greater number of responders (i.e. patients) and therefore conjoint analysis techniques could be used⁴⁵².

References

1. Department for Communities and Local Government. Multi-criteria analysis: a manual. 2009;
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/7612/1132618.pdf. Accessed 15 July 2013.
2. Belton V, Stewart T. *Multiple criteria decision analysis: an integrated approach*. Dordrecht: Kluwer Academic Publishers; 2002.
3. Keeney R. *Value focused thinking: a path to creative decision making*. Cambridge: Harvard University Press; 1992.
4. Beauchamp TL. Methods and principles in biomedical ethics. *Journal Of Medical Ethics*. 2003;29(5):269-274.
5. Eddy DM. Clinical decision making: from theory to practice. The individual vs society. Is there a conflict? *JAMA*. 1991;265(11):1446.
6. Eddy DM. Clinical decision making: from theory to practice. The individual vs society. Resolving the conflict. *JAMA*. 1991;265(18):2399.

7. Emanuel EJ. Justice and managed care. Four principles for the just allocation of health care resources. *The Hastings Center Report*. 2000;30(3):8-16.
8. Fleck L. Healthcare justice and rational democratic deliberation. *The American Journal Of Bioethics: AJOB*. 2001;1(2):20-21.
9. Rawls J. *A Theory of Justice*. Belknap Press of Harvard University; 1999.
10. Klein R, Maybin J. *Thinking about rationing*. The King's Fund;2012.
11. Klein R. Dimensions of rationing: who should do what? *BMJ (Clinical Research Ed)*. 1993;307(6899):309-311.
12. Klein R. Rationing in the fiscal ice age. In. *HEPL*. Vol 52010:389-396.
13. Parker R. Social administration and scarcity. In: Butterworth E, Holman R, eds. *Social welfare in modern Britain*. London: Fontana; 1975:204-212.
14. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *Journal Of Health Services Research & Policy*. 1999;4(3):174-184.
15. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *International Journal Of Technology Assessment In Health Care*. 1989;5(4):559-575.
16. Harris J. QALYfying the value of life. *Journal Of Medical Ethics*. 1987;13(3):117-123.
17. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Medical Decision Making: An International Journal Of The Society For Medical Decision Making*. 1989;9(2):142-149.
18. Beiner R. *Political judgment*. Chicago: University of Chicago Press; 1983.
19. Daniels N. Decisions about access to health care and accountability for reasonableness. *Journal Of Urban Health: Bulletin Of The New York Academy Of Medicine*. 1999;76(2):176-191.
20. Daniels N, Sabin J. The ethics of accountability in managed care reform. *Health Affairs (Project Hope)*. 1998;17(5):50-64.
21. OECD. OECD Health Data 2012 - Frequently Requested Data. 2012; <http://www.oecd.org/els/health-systems/oecdhealthdata2012-frequentlyrequesteddata.htm>. Accessed April 8, 2013.
22. Kanavos P, Ferrario A, Vondoros S, Anderson GF. Higher US Branded Drug Prices And Spending Compared To Other Countries May Stem Partly From Quick Uptake Of New Drugs. *Health Affairs*. 2013;32(4).
23. IMS. The Global Use of Medicines: Outlook Through 2015. 2011. https://www.imshealth.com/files/web/IMSH%20Institute/Reports/The%20Global%20Use%20of%20Medicines%202015/Global_Use_of_Medicines_Report_2015.pdf. Accessed 19 Sep 2013.
24. Ioannidis J. Why most clinical research is not useful. *PLoS Med*. 2016;13(6):e1002049.
25. Sculpher M, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics*. 2006;15.
26. Claxton C. OFT, VBP: QED? . *Health Economics*. 2007;16(6):545-558.
27. Kennedy I. *Appraising the value of innovation and other benefits. A short study for NICE*. July 2009.
28. Kanavos P, Manning J, Taylor D, Schurer W, Checchi K. *Implementing value-based pricing for pharmaceuticals in the UK*. London2010.
29. Cooksey D. *A review of UK Health Research Funding*. . London: HMSO;2006.

30. OFT. The pharmaceutical price regulation scheme: an OFT market study. 2007.
http://webarchive.nationalarchives.gov.uk/20140402142426/http://www.offt.gov.uk/shared_offt/reports/comp_policy/oft885.pdf. Accessed 21 May 2015.
31. DH. A new value-based approach to the pricing of branded medicines: a consultation. In: Department of Health, ed. London;2010.
32. Fogarty M. *A History of Value Theory*. Student Economic Review: Trinity College Dublin;1996.
33. Kauder E. Genesis of the Marginal Utility Theory: From Aristotle to the End of the Eighteenth Century. *The Economic Journal*. 1953;63(251).
34. Kauder E. *A History of Marginal Utility Theory*. Princeton University Press; 1965.
35. Petty W, Graunt J. *The Economic Writings of Sir William Petty, together with The Observations upon Bills of Mortality, more probably by Captain John Graunt*. Cambridge University Press; 1899.
36. Hayek F. Richard Cantillon. *Journal of Libertarian Studies*. 1931;VII(2):217-247.
37. Brewer A. Cantillon and the land theory of value. *History of Political Economy*. 1988;20(1).
38. Murphy AE. *John Law's 'Essay on a Land Bank.'* Dublin: Aeon; 1994.
39. New World Encyclopedia. John Law (economist). 2014;
[http://www.newworldencyclopedia.org/entry/John_Law_\(economist\)](http://www.newworldencyclopedia.org/entry/John_Law_(economist)). Accessed 12 September, 2016.
40. Smith A. *An Inquiry Into the Nature and Causes of the Wealth of Nations--Adam Smith*. Indianapolis: Hackett; 1993.
41. Ricardo D. On the Principles of Political Economy and Taxation. In. London: John Murray; 1817.
42. Marx K, Moore S, Aveling E. *Karl Marx: 'Capital'*. New York: Intl Publ; 1996.
43. Mill JS. *Principles of political economy, with some of their applications to social philosophy*. Reprint of 1909 London edition, Edited with an introduction by Sir William Ashley. Reprints of Economics Classics series. Fairfield, N.J.: Kelley; 1987.
44. Library of Economics and Liberty. William Stanley Jevons (1835-1882). 2008; <http://www.econlib.org/library/Enc/bios/Jevons.html>. Accessed 12 September 2016.
45. Library of Economics and Liberty. Carl Menger (1840-1921). 2008; <http://www.econlib.org/library/Enc/bios/Menger.html>.
46. Library of Economics and Liberty. Leon Walras (1834-1910). 2008; <http://www.econlib.org/library/Enc/bios/Walras.html>. Accessed 12 September 2016.
47. Walras L. *Elements of Pure Economics. Translated and annotated by William Jaffe*. London: Allen and Unwin; 1954.
48. Marshall A. Principles of Economics. In. London: Macmillan and Co; 1890.
49. Library of Economics and Liberty. Alfred Marshall (1842-1924). 2008; <http://www.econlib.org/library/Enc/bios/Marshall.html>. Accessed 12 Septemeber 2016.
50. Library of Economics and Liberty. Arthur Cecil Pigou (1877-1959). 2008; <http://www.econlib.org/library/Enc/bios/Pigou.html>. Accessed 12 September 2016.

51. Pigou A. *The Economics of Welfare*. 4th ed. London: Macmillan; 1932.
52. Howard R. *Decision analysis: applied decision theory*. Menlo Park, CA: Stanford Research Institute; 1966.
53. Raiffa H. *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley; 1968.
54. von Winterfeldt D, Edwards W. *Decision analysis and behavioral research*. Cambridge: Cambridge University Press; 1986.
55. Savage L. *The foundations of statistics*. New York: Wiley; 1954.
56. von Winterfeldt D, Fasolo B. Structuring decision problems: A case study and reflections for practitioners. 2009;199:857-866.
57. Stewart TJ. Simplified Approaches for Multicriteria Decision Making under Uncertainty. *Journal of Multi-Criteria Decision Analysis*. 1995;4(4):246-258.
58. Figueira J, Greco S, Ehrgott M. *Multiple criteria decision analysis: state of the art surveys*. New York: Springer; 2005.
59. Cooter R, Neve M, Nutton V. *History of Medicine 1988–1992*. London.
60. Wagstaff A, Van Doorslaer E. Equity in Health Care: concepts and definitions. In: Van Doorslaer E, Wagstaff A, Rutten F, eds. *Equity in the finance and delivery of health care: an international perspective*. Oxford: Oxford University Press; 1993.
61. Webster J, Lines J, Bruce J, Armstrong Schellenberg JR, Hanson K. Which delivery systems reach the poor? A review of equity of coverage of ever-treated nets, never-treated nets, and immunisation to reduce child mortality in Africa. *The Lancet Infectious Diseases*. 2005;5(11):709-717.
62. Bennett S, Chanfreau C. Approaches to rationing antiretroviral treatment: ethical and equity implications. *Bulletin Of The World Health Organization*. 2005;83(7):541-547.
63. Olsen ØE. Bridging the equity gap in maternal and child health: health systems research is needed to improve implementation. *BMJ (Clinical Research Ed)*. 2005;331(7520):844-844.
64. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness And Resource Allocation: C/E*. 2006;4:14-14.
65. Golan O, Hansen P, Kaplan G, Tal O. Health technology prioritization: Which criteria for prioritizing new technologies and what are their relative weights? *Health policy*. 2011;102(2):126-135.
66. Guindo LA, Wagner M, Baltussen R, et al. From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. *Cost Effectiveness And Resource Allocation: C/E*. 2012;10(1):9-9.
67. Rawlins M, Barnett D, Stevens A. Pharmacoeconomics: NICE's approach to decision making *British Journal of Clinical Pharmacology*. 2010;70(3):346-349.
68. Devlin NJ, Sussex J. *Incorporating Multiple Criteria in HTA: Methods and Processes*. Office of Health Economics; 2011.
69. Antoñanzas F, Terkola R, Postma M. The Value of Medicines: A Crucial but Vague Concept. *Pharmacoeconomics*. 2016.
70. McDaniels TL, Gregory RS, Fields D. Democratizing Risk Management: Successful Public Involvement in Local Water Management Decisions. *Risk Analysis: An International Journal*. 1999;19(3):497-510.

71. Bazerman M. *Judgment in Managerial Decision Making*. Fourth ed: John Wiley, New York; 1998.
72. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. *PharmacoEconomics*. 2014;32(4):345-365.
73. EMA. Benefit-risk methodology project. Work package 4 report: Benefit-risk tools and processes. 2012.
http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf. Accessed 26 July 2016.
74. Institute of Medicine. *Ranking vaccines a prioritization framework phase I: demonstration of concept and a software blueprint*. Washington DC: National Academies Press; 2012.
75. Drummond M, Bridges J, Muhlbacher A, Ijzerman M. Identification, weighting and prioritization of multiple endpoints for comparative effectiveness research: what have we learned from Germany? In: ISPOR Panel 20112011.
76. NICE. Interim Process and Methods of the Highly Specialised Technologies Programme. 2013. <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-highly-specialised-technologies-guidance/Highly-Specialised-Technologies-Interim-methods-and-process-statements.pdf>. Accessed 21 May 2015.
77. Pearman AD, Mackie PJ, May AD, Simon D. The use of multi-criteria techniques to rank highway investment proposals. In: Lockett AG, Islei G, eds. *Improving Decision Making in Organisations*. Springer; 1989.
78. DETR. A New Deal for Transport: Better for Everyone (The Government's White Paper on the Future of Transport). In: The Stationery Office; 1998.
79. Wilson EC, Peacock SJ, Ruta D. Priority setting in practice: what is the best way to compare costs and benefits? *Health Economics*. 2009;18(4):467-478.
80. Baltussen R, Youngkong S, Paolucci F, Niessen L. Multi-criteria decision analysis to prioritize health interventions: Capitalizing on first experiences. *Health Policy*. 2010;96(3):262-264.
81. Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Economics*. 2006;15(7):689-696.
82. Baltussen R, Asbroek AHA, Koolman X, Shrestha N, Bhattarai P, Niessen LW. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health Policy And Planning*. 2007;22(3):178-185.
83. Youngkong S, Teerawattananon Y, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. *Health Research Policy And Systems / Biomed Central*. 2012;10:6-6.
84. Youngkong S, Baltussen R, Tantivess S, Mohara A, Teerawattananon Y. Multicriteria decision analysis for including health interventions in the universal health coverage benefit package in Thailand. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2012;15(6):961-970.
85. Phillips LD, Fasolo B, Zafiroopoulos N, Beyer A. Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug discovery today: technologies*. 2011;8(1):e3-e10.
86. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using

- multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf.* 2007;16(Suppl 1):S2-S15.
87. Walker S, Cone M. *Benefit–Risk Assessment: Summary Report of the workshop on The Development of a Model for Benefit–Risk Assessment of Medicines Based on Multi-Criteria Decision Analysis.* . CMR International Institute for Regulatory Science;2004.
 88. Health England. *Prioritising investments in preventative health.* 2009.
 89. Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. *Multi-Criteria Decision Analysis to Value Orphan Medicines.* Office of Health Economics;2013.
 90. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Evidence and Value: Impact on DEcisionMaking--the EVIDEM framework and potential applications. *BMC Health Services Research.* 2008;8:270.
 91. Thokala P. *Multi criteria decision analysis for health technology assessment: report by the decision support unit.* Sheffield: University of Sheffield;2011.
 92. Hammond J, Keeney R, Raiffa H. *Smart choices: a practical guide to making better decision.* Cambridge: Harvard University Press; 1999.
 93. Bana e Costa C, Nunes da Silva F, Vansnick J-C. Conflict dissolution in the public sector: a case study. *European Journal of Operational Research.* 2000.
 94. Phillips L, Phillips M. Facilitated Work Groups: Theory and Practice. *The Journal of the Operational Research Society.* 1993;44(6).
 95. Phillips LD. Decision Conferencing. In: Edwards W, Miles Jr RF, von Winterfeldt D, eds. *Advances in Decision Analysis: From Foundations to Applications.* Cambridge University Press; 2007.
 96. van Wyk B. Research design and methods: Part I.
https://www.google.gr/url?sa=t&amrct=j&amq=&amescr=s&amsource=web&amcd=7&amved=0ahUKEwiT3tbV5dnOAhXsBsAKHVO7DqgQFghCMAY&amurl=https%3A%2F%2Fwww.uwc.ac.za%2Fstudents%2Fpostgraduate%2Fdocuments%2Fresearch_and_design_i.pdf&amug=AfQjCNEiQkvCdI3grtgEnQWE14yoVMDMg&amsig2=X17AJzwjCAvjzVisXCFuwg&amcad=rja. Accessed 21 May 2015.
 97. Creswell J. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches.* Second ed: Sage Publications; 2003.
 98. Banta H, Behney C, Andrudid D. *Assessing the efficacy and safety of medical technologies.* Washington: Office of Technology Assessment;1978.
 99. Luce B, Drummond M, Jonsson B, et al. EBM, HTA, and CER: Clearing the Confusion. *Milbank Quarterly.* 2010;88(2):256–276.
 100. Taylor R, Drummond M, Salkeld G, Sullivan S. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. *BMJ.* 2004;329:972–975.
 101. Kanavos P, Nicod E, van den Aardweg S, Pomedli S. The impact of health technology assessments: an international comparison. *Euro Observer.* 2010b;12(4).
 102. Nicod E. *Major Review Document: A Critical Appraisal of Coverage and Resource Allocation Decisions through the use of Health Technology Assessment: Evidence from 4 countries.* London School of Economics;2012.
 103. Bombardier C, Maetzel A. Pharmacoeconomic evaluation of new treatments: efficacy versus effectiveness studies? . *Ann Rheum Dis.* 1999;58(Suppl 1):I82-85.

104. Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. *Lancet*. 2006;368:1470-1472.
105. Falissard B, Zylberman M, Cucherat M, Izard V, Meyer F, al. e. Real Medical Benefit Assessed by Indirect Comparison. *Thérapie*. 2009;64(3):229-232.
106. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;20:3105-3124.
107. Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the economic evaluation of health care Programmes. In. Oxford: Oxford University Press; 1987.
108. Drummond M. Basic types of economic evaluation. In: Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G, eds. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2005.
109. Besley T. Welfare Economics and Public Choice. 2002. <http://econ.lse.ac.uk/~tbesley/papers/welfpub.pdf>.
110. Mishan E. *Welfare economics: an assessment*. 1969.
111. McGuire A. Theoretical Concepts in the Economic Evaluation of Health Care. In: Drummond MF, McGuire A, eds. *Economic Evaluation in Health Care*. New York: Oxford University Press; 2001:1-21.
112. Ng Y. *Welfare Economics*. Macmillan; 1979.
113. Samuelson P. *Foundations of Economic Analysis*. Harvard University Press; 1947.
114. Barr N. *Economics of the Welfare State*. Fifth ed: OUP Oxford; 2012.
115. Drummond M, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Basic types of economic evaluation. In: Drummond MF, ed. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2005:7-26.
116. Richardson J. Cost- utility analysis: what should be measured? *Soc Sci Med*. 1994;39(1):7-21.
117. George B, Harris A, Mitchell A. Cost-Effectiveness Analysis and the Consistency of Decision Making: Evidence from Pharmaceutical Reimbursement in Australia (1991 to 1996). *Pharmacoeconomics*. 2001;19(11):1103-1109.
118. Schlander M. Measures of efficiency in healthcare: QALMs about QALYs? *Z Evid Fortbild Qual Gesundh wesen (ZEFQ)*. 2010;104:209-226.
119. Kanavos P, Ferrario A. *Manged entry agreements for pharmaceuticals: The European experience*. DG Enterprise; 2013.
120. Towse A, Garrison L. Can't Get No Satisfaction? Will Pay for Performance Help? *Pharmacoeconomics*. 2010;28(2):93-102.
121. Gold MR, Siegel JE, L.B. R, al. e. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.
122. Ghislandi S, Apolone G, Garattini L, Ghislandi I. Is EQ-5D a Valid Measure of HRQoL in Patients with Movement Disorders? A Comparison with SF-36 and FIM questionnaires. *The European Journal of Health Economics*. 2002;3:125-130.
123. Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26(9):769-779.
124. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*. 2004;13(5):437-452.

125. NICE. *Guidance On the Use of Orlistat for the Treatment of Obesity in Adults. Technology Appraisal Guidance No. 22*. National Institute for Health and Clinical Excellence;2002.
126. NICE. *Health Select Committee Inquiry: Submission of Supplementary Evidence*. National Institute for Health and Clinical Excellence;2002b.
127. NICE. *Guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence;2008.
128. Towse A. What is NICE's threshold? An external view. In: Towse A, Pritchard C, Devlin N, eds. *Cost Effectiveness Thresholds: Economic and Ethical Issues*. London: King's Fund/Office for Health Economics; 2002.
129. Rawlins M, Culyer A. National Institute of Clinical Excellence and its value judgments. *BMJ*. 2004;(329):224–227.
130. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to Pay for a Quality-adjusted Life Year : In Search of a Standard. *Medical Decision Making*. 2000;20.
131. Dakin HA, Devlin NJ, Odeyemi IAO. “Yes”, “No” or “Yes, but”? Multinomial modelling of NICE decision-making. *Health Policy*. 2006;77(3):352-367.
132. Kanavos P, Casson M, Nicod E. *A new value-based approach to the pricing of branded medicines: a consultation*. LSE Health;2011b.
133. Rocchi A, Menon D, Verma S, Miller E. The Role of Economic Evidence in Canadian Oncology Reimbursement Decision-Making: To Lambda and Beyond. *Value in Health*. 2008;11(4):771-783.
134. Simoens S, Doooms M. How much is the life of a cancer patient worth? A pharmaco-economic perspective. *Journal Of Clinical Pharmacy And Therapeutics*. 2011;36(3):249-256.
135. Bredesen G. Organization of health care systems and financing of pharmaceuticals in Norway, including cost containments. *Pharmaceut Finance Econ Pol*. 2003;12:351-360.
136. Carlsson P. Health technology assessment and priority setting for health policy in Sweden. *International Journal Of Technology Assessment In Health Care*. 2004;20(1):44-54.
137. Leplège A, Hunt S. The problem of quality of life in medicine. *JAMA*. 1997;278(1):47-50.
138. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Social Science and Medicine (1982)*. 1989;28(4):299-308.
139. NICE. *Value Based Pricing: Response to the Department of Health consultation*. National Institute for Health and Clinical Excellence;2011.
140. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*. 2004;13:437-452.
141. Earnshaw J, Lewis G. NICE Guide to the Methods of Technology Appraisal - Pharmaceutical industry perspective. *Pharmacoeconomics*. 2008;26(9):725-727.
142. Harris AH, Hill SR, Chin G, Li JJ, Walkom E. The Role of Value for Money in Public Insurance Coverage Decisions for Drugs in Australia: A Retrospective Analysis 1994-2004. *Med Decis Making*. 2008;28.
143. Le Pen C, Priol G, Lilliu H. What criteria for pharmaceuticals reimbursement? *Eur J Health Econ*. 2003;4:30-36.

144. Pronk MH, Bonsel GJ. Out-patient drug policy by clinical assessment rather than financial constraints? The gate-keeping function of the out-patient drug reimbursement system in The Netherlands *Eur J Health Econ*. 2004;5(3):274-277.
145. Sussex J, Towse A, Devlin N. Operationalizing Value- Based Pricing of Medicines. *PharmacoEconomics*. 2013;31(1):1-10.
146. NICE. *Guide to the technology appraisal process*. London: National Institute for Health and Clinical Excellence;2004.
147. NICE. *Department of Health Selection Criteria*. London: National Institute for Health and Clinical Excellence;2006.
148. Shah KK, Cookson R, Culyer AJ, Littlejohns P. NICE's social value judgements about equity in health and health care. *Health Economics Policy and Law*. 2013;8:145-165.
149. McGuire A, Raikou M, Kanavos P. Pricing pharmaceuticals: Value based pricing in what sense. *Eurohealth*. 2008;14(2):3-6.
150. Sloan FA, Hsieh C-R. Introduction. In: Sloan FA, Hsieh C-R, eds. *Pharmaceutical Innovation: Incentives, Competition, and Cost-Benefit Analysis in International Perspective*. New York: Cambridge University Press; 2007.
151. Dakin HA, Devlin NJ, Odeyemi IAO. “Yes”, “No” or “Yes, but”? Multinomial modelling of NICE decision-making. *Health Policy*. 2006;77:352-367.
152. Daniels N. Decisions about access to health care and accountability for reasonableness. *J Urban Health*. 1999;76:176-191.
153. Longson C, Littlejohns P. *Update Report on the Application of the 'End-of-Life' Supplementary Advice in Health Technology Appraisals*. London: NICE;2009.
154. DH. A new value-based approach to the pricing of branded medicines: government response to consultation. In: Department of Health, ed. London;2011.
155. Morgan SG, McMahon M, Mitton C, et al. Centralized Drug Review Processes In Australia, Canada, New Zealand, And The United Kingdom. *Health Affairs*. 2006;25(2):337-347.
156. Annis AH, Guh D, Wang X-H. A Dog's Breakfast: Prescription Drug Coverage Varies Widely Across Canada. *Medical Care*. 2001;39(4):315-326.
157. Anell A. Priority setting for pharmaceuticals. The use of health economic evidence by reimbursement and clinical guidance committees. *Eur J Health Econ*. 2004;5:28-35.
158. Gregoire J, MacNeil P, Skilton K, et al. Inter-provincial variation in government drug formularies. *Can J Public Health*. 2001;92:307-312.
159. Schwarzer R, Siebert U. Methods, procedures, and contextual characteristics of health technology assessment and health policy decision making: comparison of health technology assessment agencies in Germany, United Kingdom, France, and Sweden. *International Journal Of Technology Assessment In Health Care*. 2009;25(3):305-314.
160. Banta D. The development of health technology assessment. *Health Policy*. 2003;63(2):121.
161. Cairns J. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute of Clinical Excellence. *Health Policy*. 2006;76:134-143.

162. Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using Effectiveness and Cost-effectiveness to Make Drug Coverage Decisions : A Comparison of Britain, Australia, and Canada. *JAMA*. 2009;302(13):1437-1443.
163. Busse R, Orvain J, Velasco M, et al. Best practice in undertaking and reporting health technology assessments - Working group 4 report. *International Journal of Technology Assessment in Health care*. 2002;18(2):361-422.
164. Sorenson C, Drummond M, Kanavos P. *Ensuring Value for Money in Health Care: the role of HTA in the European Union*. Copenhagen: European Observatory on Health Systems and Policies;2008b.
165. Teng F, Mitton C, Mackenzie J. Priority setting in the provincial health services authority: survey of key decision makers. *BMC Health Serv Res*. 2007;7:84.
166. Dionne F, Mitton C, Smith N, Donaldson C. Decision maker views on priority setting in the Vancouver Island Health Authority. *Cost Eff Resour Alloc*. 2008;6:13.
167. Mitton C, McMahon M, Morgan S, Gibson J. Centralized drug review processes: are they fair? *Soc Sci Med*. 2006;63:200-211.
168. Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Performances and Value Trade-Offs*. New York: Wiley; 1976.
169. Neumann PJ, Cohen JT. Measuring the Value of Prescription Drugs. *The New England Journal Of Medicine*. 2015;373(27):2595-2597.
170. Anderson JL, Heidenreich PA, Barnett PG, et al. ACC/AHA statement on cost/value methodology in clinical practice guidelines and performance measures: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and Task Force on Practice Guidelines. *Journal Of The American College Of Cardiology*. 2014;63(21):2304-2322.
171. Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2015;33(23):2563-2577.
172. Cherny NI, Sullivan R, Dafni U, et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). *Annals Of Oncology: Official Journal Of The European Society For Medical Oncology / ESMO*. 2015;26(8):1547-1573.
173. Pearson S. *Framework for Payer Assessment of the Value of New Technologies: A US Approach*. Office of Health Economics;2015.
174. Bach P. DrugAbacus App. 2015; <http://app.drugabacus.org/abacus-mskcc>. Accessed June, 2016.
175. NCCN. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) with NCCN Evidence Blocks™. 2015; <http://www.nccn.org/evidenceblocks/>. Accessed 23 June, 2016.
176. Working Group on Mechanism of Coordinated Access to Orphan Medicinal Products. *Process on Corporate Social Responsibility in the Field of Pharmaceuticals Platform on Access to Medicines in Europe. Final Report*. Brussels: European Commission;2013.

177. Angelis A, Kanavos P. Critique of the American Society of Clinical Oncology Value Assessment Framework for Cancer Treatments: Putting Methodologic Robustness First. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2016.
178. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer A, Newhouse J, eds. *Handbook of health economics*. Amsterdam;2000.
179. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E. Toward a consensus on the QALY. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2009;12 Suppl 1:S31-S35.
180. Weinstein MC, Torrance G, McGuire A. QALYs: the Basics. *Value in Health*. 2009;12(Supplement s1):s5-s9.
181. Kahneman D. New Challenges to the Rationality Assumption. *Journal of Institutional and Theoretical Economics*. 1994;150(1):18-36.
182. Dolan P, Kahneman D. Interpretations Of Utility And Their Implications For The Valuation Of Health. *Economic Journal*. 2008;118(525):215-234.
183. Normand C. Measuring outcomes in palliative care: limitations of QALYs and the road to PaLYs. *Journal Of Pain And Symptom Management*. 2009;38(1):27-31.
184. Garau M, Shah KK, Mason AR, Wang Q, Towse A, Drummond MF. Using QALYs in Cancer: A Review of the Methodological Limitations. *Pharmacoeconomics*. 2011;29(8):673-685.
185. Tordrup D, Mossman J, Kanavos P. Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented? *International Journal Of Technology Assessment In Health Care*. 2014;30(1):10-19.
186. Zwart-van Rijkom JEF, Leufkens HGM, Busschbach JJV, Broekmans AW, Rutten FFH. Differences in Attitudes, Knowledge and Use of Economic Evaluations in Decision-Making in The Netherlands: The Dutch Results from the EUROMET Project. *Pharmacoeconomics*. 2000;18(2):149-160.
187. NICE. Appraising life-extending, end of life treatments. 2009. <https://www.nice.org.uk/guidance/gid-tag387/resources/appraising-life-extending-end-of-life-treatments-paper2>. Accessed 14 July 2016.
188. Claxton K. OFT, VBP: QED? *Health Economics*. 2007;16(6):545-558.
189. Towse A. If it ain't broke, don't price fix it: the OFT and the PPRS. *Health Economics*. 2007;16(7):653-665.
190. NICE. Consultation paper: value based assessment of health technologies. In. London: National Institute for Health and Care Excellence; 2014.
191. Devlin N, Sussex J. Incorporating multiple criteria in HTA: methods and processes. In. London: Office of Health Economics; 2011.
192. Kanavos P, Angelis A. Multiple criteria decision analysis for value based assessment of new medical technologies: a conceptual framework. In. *The LSE Health Working Paper Series in Health Policy and Economics*: London School of Economics and Political Science; 2013.
193. Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. *Multi-criteria decision analysis to value orphan medicines*. London: Office of Health Economics;2013.
194. Radaelli G, Lettieri E, Masella C, Merlino L, Strada A, Tringali M. Implementation of EUnetHTA core Model® in Lombardia: the VTS

- framework. *International Journal Of Technology Assessment In Health Care*. 2014;30(1):105-112.
195. Dolan JG. Multi-criteria clinical decision support: A primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. *The Patient*. 2010;3(4):229-248.
 196. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2012;15(8):1172-1181.
 197. Diaby V, Goeree R. How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: a step-by-step guide. *Expert Review Of Pharmacoeconomics and Outcomes Research*. 2014;14(1):81-99.
 198. Roy B. Critères multiples et modélisation des préférences : l'apport des relations de surclassement. *Revue d'Economie Politique*. 1974;1:1-44.
 199. Goodwin P, Wright G. *Decision Analysis for Management Judgement*. second ed. Chichester: John Wiley & Sons; 1997.
 200. Marsh K, Ijzerman M, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making-Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2016;19(2):125-137.
 201. Angelis A, Kanavos P. Using MCDA to assess value of new medical technologies. In. *LSE Health Working Paper Series in Health Policy and Economics*: London School of Economics and Political Science; 2013.
 202. Keeney RL. Common Mistakes In Making Value Trade-Offs. *Operations Research*. 2002;50(6):935.
 203. Bana E Costa CA, De Corte J-M, Vansnick J-C. MACBETH. *International Journal of Information Technology & Decision Making*. 2012;11(2):359-387.
 204. Franco L, Montibeller G. Problem Structuring for Multicriteria Decision Analysis Interventions analysis interventions. *Wiley Encyclopedia of Operations Research and Management Science*. 2010.
 205. Keeney R, Raiffa H. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge: Cambridge University Press; 1993.
 206. Keeney RL, Gregory RS. Selecting Attributes to Measure the Achievement of Objectives. *Operations Research*. 2005;53(1):1-11.
 207. Claxton K. Should multi-criteria decision analysis (MCDA) replace cost effectiveness analysis (CEA) for evaluation of health care coverage decisions? Paper presented at: ISPOR 16th Annual European Congress;2013; Dublin.
 208. Broekhuizen H, Groothuis-Oudshoorn C, Til J, Hummel J, Ijzerman M. A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions. *PharmacoEconomics*. 2015;33(5):445-455.
 209. Brouwer W, Van Exel J, Baker R, Donaldson C. The new myth. *PharmacoEconomics*. 2008;26(1):1-4.
 210. Wouters OJ, Naci H, Samani NJ. QALYs in cost-effectiveness analysis: an overview for cardiologists. *Heart (British Cardiac Society)*. 2015;101(23):1868-1873.
 211. Goldman D, Lakdawalla D, Philipson TJ, Yin W. Valuing health technologies at nice: recommendations for improved incorporation of treatment value in HTA. *Health Economics*. 2010;19(10):1109-1116.
 212. Runge C. Separating the wheat from the chaff. In. Vol 132012:121.

213. Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. *Health Policy*. 2016;120(1):35-45.
214. Anell A. Priority setting for pharmaceuticals. The use of health economic evidence by reimbursement and clinical guidance committees. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2004;5(1):28-35.
215. Kanavos P, Nicod E, van der Aardweg S, Pomedli S. The impact of health technology assessments: an international comparison. *The Health Policy Bulletin of the European Observatory on Health Systems and Policies*. 2010;12(4).
216. Morgan SG, McMahon M, Mitton C, et al. Centralized drug review processes in Australia, Canada, New Zealand, and the United kingdom. *Health Affairs (Project Hope)*. 2006;25(2):337-347.
217. Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: A comparative analysis of five countries and implications for coverage decisions. *Health Policy*. 2012;108(2/3):167-177.
218. Dionne F, Mitton C, Smith N, Donaldson C. Decision maker views on priority setting in the Vancouver Island Health Authority. *Cost Effectiveness And Resource Allocation: C/E*. 2008;6:13-13.
219. Mitton CR, McMahon M, Morgan S, Gibson J. Centralized drug review processes: are they fair? *Social Science and Medicine (1982)*. 2006;63(1):200-211.
220. Teng F, Mitton C, Mackenzie J. Priority setting in the provincial health services authority: survey of key decision makers. *BMC Health Services Research*. 2007;7:84-84.
221. Kanavos P. *The future of Health Technology Assessment: Evidence from Europe and the Americas and an agenda for policy action*. Vol forthcoming. Brussels: European Observatory; 2017.
222. Advance-HTA. Advance-HTA webpage. 2013; <http://www.advance-hta.eu/>. Accessed June, 2016.
223. Hutton J, McGrath C, Frybourg J-M, Tremblay M, Bramley-Harker E, Henshall C. Framework for describing and classifying decision- making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care*. 2006;22(1):10-18.
224. Rogowski WH, Hartz SC, John JH. Clearing up the hazy road from bench to bedside: a framework for integrating the fourth hurdle into translational medicine. *BMC Health Services Research*. 2008;8:194-194.
225. Fischer KE. Decision- making in healthcare: a practical application of partial least square path modelling to coverage of newborn screening programmes. *BMC medical informatics and decision making*. 2012;12:83.
226. Centre for Reviews and Dissemination. Systematic reviews : CRD's guidance for undertaking reviews in health care. In: [3rd ed.]. ed.: University of York; 2009: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Accessed 21 June 2016.
227. Advance-HTA. Personal communication with experts. In: Advance-HTA Project; 2016.
228. AHTAPol. *Guidelines for conducting Health Technology Assessment (HTA)*. Warsaw: Agency for Health Technology Assessment;2009.

229. AIFA. The Italian Medicines Agency.
<http://www.agenziafarmaco.com/en/node/4111>. Accessed June, 2016.
230. Antónanzas F. Challenges to Achieving Value in Drug Spending in a Decentralized Country: The Spanish Case. *Value in Health*. 2003;6:S52-S63.
231. Banta D, de Wit A. Public health services and cost-effectiveness analysis. *Annual Review Of Public Health*. 2008;29:383-397.
232. Banta D, Oortwijn WJ. The Netherlands. *International Journal Of Technology Assessment In Health Care*. 2009;25 Suppl 1:143-147.
233. Bell KJL, Irwig L, March LM, Hayen A, Macaskill P, Craig JC. Should response rules be used to decide continued subsidy of very expensive drugs? A checklist for decision makers. *Pharmacoepidemiology And Drug Safety*. 2010;19(1):99-105.
234. Boersma C, Broere A, Postma MJ. Quantification of the potential impact of cost-effectiveness thresholds on dutch drug expenditures using retrospective analysis. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2010;13(6):853-856.
235. Briggs AH, Levy AR. Pharmacoeconomics and Pharmacoepidemiology: Curious bedfellows or a match made in heaven? In. Vol 242006:1079-1086.
236. Bryan S, Williams I, McIver S. Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. *Health Economics*. 2007;16(2):179-193.
237. Buxton M. Implications of the appraisal function of the National Institute for Clinical Excellence (NICE). *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2001;4(3):212-216.
238. Calltorp J. Priority setting in health policy in Sweden and a comparison with Norway. *Health Policy (Amsterdam, Netherlands)*. 1999;50(1-2):1-22.
239. Capri S, Ceci A, Terranova L, Merlo F, Mantovani L. Guidelines for economic evaluations in Italy: Recommendations from the Italian Group of Pharmacoeconomic Studies. *Drug Information Journal*. 2001;35:189-201.
240. Catalá-López F, Alonso-Arroyo A, Aleixandre-Benavent R, et al. Coauthorship and institutional collaborations on cost-effectiveness analyses: a systematic network analysis. *Plos One*. 2012;7(5):e38012-e38012.
241. Chalkidou K. Evidence and values: paying for end-of-life drugs in the British NHS. *Health Economics, Policy, And Law*. 2012;7(4):393-409.
242. Cleemput I, Franken M, Koopmanschap M, le Polain M. European drug reimbursement systems' legitimacy: five-country comparison and policy tool. *International Journal Of Technology Assessment In Health Care*. 2012;28(4):358-366.
243. Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA*. 2009;302(13):1437-1443.
244. Cookson R, McDaid D, Maynard A. Wrong SIGN, NICE mess: is national guidance distorting allocation of resources? *BMJ (Clinical Research Ed)*. 2001;323(7315):743-745.
245. Corbacho B, Pinto-Prades JL. Health economic decision-making: a comparison between UK and Spain. *British Medical Bulletin*. 2012;103(1):5-20.

246. CVZ. Guidelines for pharmacoeconomic research, updated version. 2006. <https://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/publicaties/publications-in-english/2006/0604-guidelines-for-pharmacoeconomic-research/0604-guidelines-for-pharmacoeconomic-research/Guidelines+for+pharmacoeconomic+research.pdf>. Accessed 21 June 2016.
247. DH. Impact Assessment of Proposal for a Cancer Drug Fund. 2010. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_120930.pdf. Accessed 21 June 2015.
248. Dirksen CD, Utens CM, Joore MA, et al. Integrating evidence on patient preferences in healthcare policy decisions: protocol of the patient-VIP study. *Implementation Science: IS*. 2013;8:64-64.
249. Drummond M, Jönsson B, Rutten F, Stargardt T. Reimbursement of pharmaceuticals: reference pricing versus health technology assessment. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2011;12(3):263-271.
250. Falissard B, Izard V, Xerri B, Bouvenot G, Meyer F, Degos L. Relative effectiveness assessment of listed drugs (REAL): a new method for an early comparison of the effectiveness of approved health technologies. *International Journal Of Technology Assessment In Health Care*. 2010;26(1):124-130.
251. Favaretti C, Cicchetti A, Guarrera G, Marchetti M, Ricciardi W. Health technology assessment in Italy. *International Journal Of Technology Assessment In Health Care*. 2009;25 Suppl 1:127-133.
252. Folino-Gallo P, Montilla S, Bruzzzone M, Martini N. Pricing and reimbursement of pharmaceuticals in Italy. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2008;9(3):305-310.
253. Franken M, Nilsson F, Sandmann F, Boer A, Koopmanschap M. Unravelling Drug Reimbursement Outcomes: A Comparative Study of the Role of Pharmacoeconomic Evidence in Dutch and Swedish Reimbursement Decision Making. *PharmacoEconomics*. 2013;31(9):781-797.
254. Freemantle N. Biostatistical aspects for the use of evidence based medicine in health technology assessment. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2008;9 Suppl 1:31-43.
255. Gafni A, Birch S. NICE Methodological Guidelines and Decision Making in the National Health Service in England and Wales. *PharmacoEconomics*. 2003;21(3):149-157.
256. Gridchyna I, Aulois-Griot M, Maurain C, Bégaud B. How innovative are pharmaceutical innovations?: the case of medicines financed through add-on payments outside of the French DRG-based hospital payment system. *Health Policy (Amsterdam, Netherlands)*. 2012;104(1):69-75.
257. HAS. General method for assessing health technologies. 2007. http://www.has-sante.fr/portail/upload/docs/application/pdf/general_method_eval techno.pdf. Accessed 22 June 2016.
258. HAS. Choices in Methods for Economic Evaluation. 2012. http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf. Accessed 22 June 2016.

259. HAS. La Commission évaluation économique et de santé publique (CEESP). 2012. http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-11/quest-ce_que_la_ceesp-brochure2clics_2012-11-22_10-13-34_631.pdf. Accessed 23 June 2016.
260. HAS. Évaluation médico-économique des produits de santé. 2013; http://www.has-sante.fr/portail/jcms/r_1502595/fr/evaluation-medico-economique-des-produits-de-sante. Accessed June, 2016.
261. Hjalte K, Hjelmgren J, Johansson F, Persson U. *Betalningsviljan for ett kvalitetsjusterat levnadsar - en pilotstudie*. Institutet for Halsoekonomi;2005.
262. Hoyle M. Historical lifetimes of drugs in England: application to value of information and cost-effectiveness analyses. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2010;13(8):885-892.
263. Hughes-Wilson W, Palma A, Schuurman A, Simoens S. Paying for the Orphan Drug System: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? *Orphanet Journal Of Rare Diseases*. 2012;7:74-74.
264. IQWiG. General Methods for the Assessment of the Relation of Benefits to Costs. 2009. http://www.ispor.org/peguidelines/source/Germany_AssessmentoftheRelationofBenefitstoCosts_En.pdf. Accessed 22 June 2016.
265. IQWiG. General Methods. 2011. https://www.iqwig.de/download/General_Methods_4-0.pdf.
266. ISPOR. ISPOR Global Health Care Systems Road Map - Spain. 2009; <http://www.ispor.org/htaroadmaps/spain.asp>. Accessed 23 June, 2016.
267. ISPOR. Pharmacoeconomic guidelines around the world. 2014;2016(22 June). <http://www.ispor.org/peguidelines/index.asp>.
268. Jena AB, Philipson TJ. Endogenous cost-effectiveness analysis and health care technology adoption. *Journal Of Health Economics*. 2013;32(1):172-180.
269. Jönsson B. Economic evaluation for pharmaceuticals in Germany. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2007;8 Suppl 1:S1-S2.
270. Kaltenthaler E, Boland A, Carroll C, Dickson R, Fitzgerald P, Papaioannou D. Evidence Review Group approaches to the critical appraisal of manufacturer submissions for the NICE STA process: a mapping study and thematic analysis. *Health Technology Assessment (Winchester, England)*. 2011;15(22):1.
271. Kenny N, Joffres C. An ethical analysis of international health priority-setting. *Health Care Analysis: HCA: Journal Of Health Philosophy And Policy*. 2008;16(2):145-160.
272. Kirkdale R, Krell J, Hanlon Brown C, Tuthill M, Waxman J. The cost of a QALY. *QJM: An International Journal of Medicine*. 2010;103(9):715-720.
273. Kleijnen S, George E, Goulden S, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2012;15(6):954-960.
274. Kolasa K, Schubert S, Manca A, Hermanowski T. A review of Health Technology Assessment (HTA) recommendations for drug therapies issued between 2007 and 2009 and their impact on policymaking processes in Poland. *Health Policy*. 2011;102(2/3):145-151.

275. Le Pen C, Priol G, Lilliu H. What criteria for pharmaceuticals reimbursement? An empirical analysis of the evaluation of "medical service rendered" by reimbursable drugs in France. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2003;4(1):30-36.
276. Le Pen C. Pharmaceutical economy and the economic assessment of drugs in France. *Social Science & Medicine (1982)*. 1997;45(4):635-643.
277. López-Bastida J, Oliva J, Antoñanzas F, et al. Spanish recommendations on economic evaluation of health technologies. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2010;11(5):513-520.
278. Manchikanti L, Falco FJE, Boswell MV, Hirsch JA. Facts, fallacies, and politics of comparative effectiveness research: Part 2 - implications for interventional pain management. *Pain Physician*. 2010;13(1):E55-E79.
279. Mangiapane S, Velasco Garrido M. Use of Surrogate end points in HTA. *GMS Health Technology Assessment*. 2009;5:Doc12-Doc12.
280. Mapelli V, Lucioni C. Spending on pharmaceuticals in Italy: macro constraints with local autonomy. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2003;6 Suppl 1:S31-S45.
281. Mason AR, Drummond MF. Public funding of new cancer drugs: Is NICE getting nastier? *European Journal Of Cancer (Oxford, England: 1990)*. 2009;45(7):1188-1192.
282. Mauskopf JA, Sullivan SD, Annemans L, et al. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices--budget impact analysis. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2007;10(5):336-347.
283. Mauskopf J, Walter J, Birt J, Bowman L, Copley-Merriman C, Drummond M. Differences among formulary submission guidelines: implications for health technology assessment. *International Journal Of Technology Assessment In Health Care*. 2011;27(3):261-270.
284. McGhan WF, Al M, Doshi JA, Kamae I, Marx SE, Rindress D. The ISPOR Good Practices for Quality Improvement of Cost-Effectiveness Research Task Force Report. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2009;12(8):1086-1099.
285. McMahon M, Morgan S, Mitton C. The Common Drug Review: A NICE start for Canada? *Health Policy*. 2006;77(3):339-351.
286. Milne R, Clegg A, Stevens A. HTA responses and the classic HTA report. *Journal Of Public Health Medicine*. 2003;25(2):102-106.
287. NICE. Social Value Judgements: Principles for the development of NICE guidance. 2008. <https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf>. Accessed 23 June 2016.
288. NICE. *The guidelines manual*. National Institute for Health and Clinical Excellence;2009.
289. NICE. Guide to the methods of technology appraisal 2013. 2013. <https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf>. Accessed 21 May 2015.

290. Nizankowski R, Wilk N. From idealistic rookies to a regional leader: The history of health technology assessment in Poland. *International Journal of Technology Assessment in Health Care*. 2009;25(S1):156-162.
291. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Standardisation of costs: The Dutch Manual for Costing in Economic Evaluations. *PharmacoEconomics*. 2002;20(7):443-454.
292. Ozieranski P, McKee M, King L. The politics of health technology assessment in Poland. *Health policy*. 2012;108(2-3):178-193.
293. Persson U, Hjelmgren J. Healso-och sjukvarden behover kunskap om hur befolkningen varderar halsan. *Lakartidningen*. 2003;100(43):3436-3437.
294. Persson U, Willis M, Ödegaard K. A case study of ex ante , value- based price and reimbursement decision- making: TLV and rimonabant in Sweden. *Eur J Health Econ*. 2010;11(2):195-203.
295. Raftery J. Review of NICE's recommendations, 1999- 2005. *British Medical Journal*. 2006;332(7552):1266-1268.
296. Rawlins M, Culyer A. National institute for clinical excellence and its value judgments. *British Medical Journal*. 2004;329(7459):224-227.
297. Riedel R, Repschläger U, Griebenow R, Breilkopf S, Schmidt S, Guhl A. International standards for health economic evaluation with a focus on the German approach. *Journal Of Clinical Pharmacy And Therapeutics*. 2013;38(4):277-285.
298. Rodríguez-Monguío R, Villar FA. Healthcare Rationing in Spain: Framework, Descriptive Analysis and Consequences. *PharmacoEconomics*. 2006;24(6):537-548.
299. Rodriguez JM, Paz S, Lizan L, Gonzalez P. The use of quality-adjusted life-years in the economic evaluation of health technologies in Spain: a review of the 1990-2009 literature. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2011;14(4):458-464.
300. Sabik LM, Lie RK. Priority setting in health care: Lessons from the experiences of eight countries. *International Journal For Equity In Health*. 2008;7:4-4.
301. Schlander M. Is NICE infallible? A qualitative study of its assessment of treatments for attention-deficit/hyperactivity disorder (ADHD). *Current Medical Research And Opinion*. 2008;24(2):515-535.
302. Schubert F. Health technology assessment. The pharmaceutical industry perspective. *International Journal Of Technology Assessment In Health Care*. 2002;18(2):184-191.
303. Sheldon TA, Cullum N, Dawson D, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. *BMJ (Clinical Research Ed)*. 2004;329(7473):999-999.
304. Simoens S. Use of economic evaluation in decision making: evidence and recommendations for improvement. *Drugs*. 2010;70(15):1917-1926.
305. Sjögren E. Deciding subsidy for pharmaceuticals based on ambiguous evidence. *Journal Of Health Organization And Management*. 2008;22(4):368-383.
306. Stewart A, Aubrey P, Belsey J. Addressing the health technology assessment of biosimilar pharmaceuticals. *Current Medical Research And Opinion*. 2010;26(9):2119-2126.

307. Stolk EA, de Bont A, van Halteren AR, Bijlmer RJ, Poley MJ. Role of health technology assessment in shaping the benefits package in The Netherlands. *Expert Review Of Pharmacoeconomics & Outcomes Research*. 2009;9(1):85-94.
308. Stoykova B, Drummond M, Barbieri M, Kleijnen J. The lag between effectiveness and cost-effectiveness evidence of new drugs. Implications for decision-making in health care. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2003;4(4):313-318.
309. Sun F, Schoelles K. *A systematic review of methods for health care technology horizon scanning*. Rockville, MD: Agency for Healthcare Research and Quality;2013.
310. The United Kingdom House of Commons Health Committee. National Institute for Clinical Excellence: Second Report of Session 2001–2002. 2002(HC 515).
<http://www.publications.parliament.uk/pa/cm200203/cmselect/cmhealth/395/39502.htm>. Accessed 22 June 2016.
311. TLV. Act (2002:160) on Pharmaceutical Benefits, etc. 2002.
<http://www.tlv.se/Upload/English/ENG-act-2002-160.pdf>. Accessed 22 June 2016.
312. TLV. Guide for companies when applying for subsidies and pricing for pharmaceutical products. 2012. <http://www.tlv.se/Upload/English/ENG-guide-for-companies.pdf>. Accessed 22 June 2016.
313. TLV. Handbook – reviewing the reimbursement status of pharmaceuticals. 2012. <http://www.tlv.se/Upload/English/ENG-handbook.pdf>. Accessed 22 June 2016.
314. Towse A, Pritchard C. National Institute for Clinical Excellence (NICE): Is Economic Appraisal Working? *PharmacoEconomics*. 2002;20(15):95-105.
315. Trowman R, Chung H, Longson C, Littlejohns P, Clark P. The National Institute for Health and Clinical Excellence and its role in assessing the value of new cancer treatments in England and Wales. *Clinical Cancer Research: An Official Journal Of The American Association For Cancer Research*. 2011;17(15):4930-4935.
316. Van Oostenbruggen MF, Jansen RB, Mur K, Kooijman H. Penny and pound wise: Pharmacoeconomics from a governmental perspective. *PharmacoEconomics*. 2005;23(3):219-226.
317. Van Wilder PB, Bormans VV, Dupont AG. Relative efficacy and effectiveness assessment of new pharmaceuticals in three EU member states: current practices and outcome agreement between Belgium, the Netherlands and France. *European Journal Of Clinical Pharmacology*. 2013;69(12):2037-2043.
318. von der Schulenburg JMG, Vauth C, Mittendorf T, Greiner W. Methods for determining cost-benefit ratios for pharmaceuticals in Germany. *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care*. 2007;8 Suppl 1:S5-S31.
319. Weill C, Banta D. Development of health technology assessment in France. *International Journal Of Technology Assessment In Health Care*. 2009;25 Suppl 1:108-111.
320. Zentner A, Velasco-Garrido M, Busse R. Methods for the comparative evaluation of pharmaceuticals. *GMS Health Technology Assessment*. 2005;1:Doc09-Doc09.

321. ZIN. Zorginstituut Nederland homepage.
<http://www.zorginstituutnederland.nl/>. Accessed 23 June, 2016.
322. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1-503, v-vi.
323. G-BA. Chapter 5: Benefit Assessment of pharmaceuticals according to s. 35a SGB V. 2015. <http://www.english.g-ba.de/downloads/17-98-3042/Chapter5-Rules-of-Procedure-G-BA.pdf>. Accessed 23 June 2016.
324. IQWiG. General Methods. 2015.
https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-2.pdf. Accessed 23 June 2016.
325. Svensson M, Nilsson F, Arnberg K. Reimbursement Decisions for Pharmaceuticals in Sweden: The Impact of Disease Severity and Cost Effectiveness. *PharmacoEconomics*. 2015;33(11):1229-1236.
326. Toumi M, Rémuzat C, El Hammi E, et al. Current process and future path for health economic assessment of pharmaceuticals in France. *Journal Of Market Access & Health Policy*. 2015;3.
327. Toumi M, Rémuzat C, El Hammi E, et al. Current process and future path for health economic assessment of pharmaceuticals in France. *Journal Of Market Access & Health Policy*. 2015;3.
328. Franken M, le Polain M, Cleemput I, Koopmanschap M. Similarities and differences between five European drug reimbursement systems. *International Journal Of Technology Assessment In Health Care*. 2012;28(4):349-357.
329. Svensson M, Nilsson FO, Arnberg K. Reimbursement Decisions for Pharmaceuticals in Sweden: The Impact of Disease Severity and Cost Effectiveness. *Pharmacoeconomics*. 2015;33(11):1229-1236.
330. WHO. Access to new medicines in Europe: technical review of policy initiatives and opportunities for collaboration and research. 2015.
<http://apps.who.int/medicinedocs/documents/s21793en/s21793en.pdf>. Accessed 23 June 2016.
331. Phillips L, Fasolo B, Zafiroopoulos N, Beyer A. Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug Discovery Today Technologies*. 2011;8(1):e1-e42.
332. Angelis A, Kanavos P. Value-Based Assessment of New Medical Technologies: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment. *Pharmacoeconomics*. 2016;34(5):435-446.
333. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2016;19(1):1-13.
334. Guyatt G. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA*. 1992;268(17).
335. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ (Clinical Research Ed)*. 1996;312(7023):71-72.
336. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Methods for the economic evaluation of health care programmes*. 4th ed: Oxford University Press; 2015.

337. Drummond M, McGuire A. *Economic Evaluation in Health Care: Merging theory with practise*. USA: Oxford University Press; 2001.
338. Murray C, Lopez A. *The Global Burden of Disease: Summary*. Harvard School of Public Health on behalf of the World Health Organisation and the World Bank; 1996.
339. Angelis A, Lange A, Kanavos P. Assessing the value of new medicines with health technology assessment: results of a systematic review and expert consultation across eight European Union countries. *European Journal of Health Economics*. 2017;In press.
340. Tervonen T, Naci H, van Valkenhoef G, et al. Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing among Statins in Primary Prevention. *Medical Decision Making: An International Journal Of The Society For Medical Decision Making*. 2015;35(7):859-871.
341. Berkeley D, Humphreys P. Structuring decision problems and the 'bias heuristic'. *Acta Psychologica*. 1982;50:201-252.
342. Buede DM. Structuring Value Attributes. *Interfaces*. 1986;16(2):52-62.
343. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiology And Drug Safety*. 2007;16 Suppl 1:S2-S15.
344. Mussen F, Salek S, Walker S, Phillips L. A quantitative approach to benefit-risk assessment of medicines - part 2: the practical application of a new model. *Pharmacoepidemiology And Drug Safety*. 2007;16 Suppl 1:S16-S41.
345. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clinical Pharmacology And Therapeutics*. 2009;85(3):241-246.
346. Mussen F, Salek S, Walker S. Benefit–Risk Appraisal of New Medicines: A Systematic Approach to Decision Making. In. Surrey, UK: John Willey; 2009.
347. FDA. Guidance for Industry: Premarketing Risk Assessment. 2005. <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126958.pdf>. Accessed 14 September 2016.
348. EMA. Report of the CHMP Working Group on Benefit–Risk Assessment Models and Methods. 2007. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf. Accessed 14 September 2016.
349. EMA. Reflection Paper on Benefit–Risk Assessment Methods in the Context of the Evaluation of Marketing Authorisation Applications of Medicinal Products for Human Use. 2008. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069634.pdf. Accessed 14 September 2016.
350. EMA. Benefit-risk methodology project. Work package 3 report: Field tests. 2011. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/09/WC500112088.pdf. Accessed 26 July 2016.
351. OHE. Response to the Department of Health’s consultation. 2011. <https://www.ohe.org/sites/default/files/old-wp/496.pdf>. Accessed 20 September 2015.

352. Miners A, Cairns J, Wailoo A. *Department of Health proposals for including Burden of Illness into Value Based Pricing: a description and critique*. NICE Decision Support Unit;2013.
353. Miners A, Cairns J, Wailoo A. *Department of Health proposals for including Wider Societal Benefits into Value Based Pricing: a description and critique*. NICE Decision Support Unit;2013.
354. NICE. Responses to TA Methods Addendum Public Consultation. 2014. <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/VBA-Consultation-Comments.pdf>. Accessed 21 May 2015.
355. Belgian Presidency. Process on corporate responsibility in the field of pharmaceuticals – Platform on access to medicines in Europe Terms of reference. In:2010: http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/tor_orphans_en.pdf. Accessed 26 July 2016.
356. Angelis A, Kanavos P. Value assessment methods and application of Multiple Criteria Decision Analysis for HTA. In. Warsaw, Poland: 1st Advance-HTA Capacity Building Workshop in Eastern Europe; 2014.
357. Kanavos P, Angelis A. Value assessment methods and application of Multiple Criteria Decision Analysis for HTA. In. Mexico City, Mexico: 1st Advance-HTA Capacity Building Workshop in Latin America; 2014.
358. Angelis A, Kanavos P. Using Multiple Criteria Decision Analysis: Conceptual Framework and Model. In. Warsaw, Poland: 2nd Advance-HTA Capacity Building Workshop in Eastern Europe; 2015.
359. Kanavos P, Angelis A. Using Multiple Criteria Decision Analysis: Conceptual Framework and Model. In. Santiago, Chile: 2nd Advance-HTA Capacity Building Workshop in Latin America; 2015.
360. Mirelman A, Mentzakis E, Kinter E, et al. Decision-making criteria among national policymakers in five countries: a discrete choice experiment eliciting relative preferences for equity and efficiency. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2012;15(3):534-539.
361. Tanios N, Wagner M, Tony M, et al. Which criteria are considered in healthcare decisions? Insights from an international survey of policy and clinical decision makers. *International Journal Of Technology Assessment In Health Care*. 2013;29(4):456-465.
362. Tromp N, Baltussen R. Mapping of multiple criteria for priority setting of health interventions: an aid for decision makers. *BMC Health Services Research*. 2012;12:454-454.
363. EUnetHTA. The HTA Core Model ®. Version 2.1. In: European Network for Health Technology Assessment: <http://www.eunethta.eu/hta-core-model>. Accessed 26 July 2016.
364. Linley WG, Hughes DA. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Economics*. 2013;22(8):948-964.
365. Hansson LF, Norheim OF, Ruyter KW. Equality, explicitness, severity, and rigidity: the Oregon plan evaluated from a Scandinavian perspective. *The Journal Of Medicine And Philosophy*. 1994;19(4):343-366.
366. European Commission. Commission Regulation (EC) No. 507/2006 on the conditional marketing authorization for medicinal products for human use.

2006. http://ec.europa.eu/health/files/eudralex/vol-1/reg_2006_507/reg_2006_507_en.pdf. Accessed 26 July 2016.
367. McCabe C, Claxton K, Tsuchiya A. Orphan drugs and the NHS: should we value rarity? *BMJ (Clinical Research Ed)*. 2005;331(7523):1016-1019.
368. European Parliament. Regulation (EC) No 141/2000 of the European Parliament and of the Council of December 16, 1999 on orphan medicinal products. 2000. http://ec.europa.eu/health/files/eudralex/vol-1/reg_2000_141/reg_2000_141_en.pdf. Accessed 26 July 2016.
369. Porter ME. What is value in health care? *The New England Journal Of Medicine*. 2010;363(26):2477-2481.
370. Biomarkers Definitions Study Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology And Therapeutics*. 2001;69(3):89-95.
371. Sullivan E. Clinical Trial Endpoints: Food and Drug Administration presentation. <http://www.fda.gov/downloads/Training/ClinicalInvestigatorTrainingCourse/UCM283378.pdf>.
372. Spahn J. Clinical trial efficacy: what does it really tell you? *The Journal Of Allergy And Clinical Immunology*. 2003;112(5 Suppl):S102-S106.
373. WHO Constitution. 1948. http://www.who.int/governance/eb/who_constitution_en.pdf. Accessed 26 July 2016.
374. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Annals Of Internal Medicine*. 1993;118(8):622-629.
375. Chen T-H, Li L, Kochen MM. A systematic review: how to choose appropriate health-related quality of life (HRQOL) measures in routine general practice? *Journal Of Zhejiang University Science B*. 2005;6(9):936-940.
376. Cherepanov D, Palta M, Fryback DG. Underlying dimensions of the five health-related quality-of-life measures used in utility assessment: evidence from the National Health Measurement Study. *Medical Care*. 2010;48(8):718-725.
377. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics (Oxford, England)*. 2000;1(1):49-67.
378. Fleming T. Design Issues in Colorectal Cancer Trials: Surrogate Endpoints and Non-Inferiority Trials. In. *Endpoints Workshop: Colorectal Cancer: Food and Drug Administration*; 2003.
379. Aronson JK, Ferner RE. Clarification of terminology in drug safety. *Drug Safety*. 2005;28(10):851-870.
380. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. 2009. http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf.
381. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clinical Pharmacology And Therapeutics*. 1981;30(2):239-245.
382. Kesselheim AS, Wang B, Avorn J. Defining "innovativeness" in drug development: a systematic review. *Clinical Pharmacology And Therapeutics*. 2013;94(3):336-348.

383. Morgan S, Lopert R, Greyson D. Toward a definition of pharmaceutical innovation. *Open Medicine: A Peer-Reviewed, Independent, Open-Access Journal*. 2008;2(1):e4-e7.
384. OHE. The many faces of innovation. 2005.
<https://www.ohe.org/publications/many-faces-innovation>. Accessed 20 September 2015.
385. Lipsey R, Carlow K. Technology policy: basic concepts. In: Edquist C, McKelvey M, eds. *Systems of Innovation: International Competitiveness*. Cheltenham: Edward Elgar; 2000.
386. Mestre-Ferrandiz J, Mordoh A, Sussex J. *The many faces of innovation: A report for the ABPI by the Office of Health Economics*. Office of Health Economics; 2012.
387. Arrow K. The economic implications of learning by doing. *Review of Economic Studies*. 1962;29(3).
388. Rosenberg N. *Inside the black box: Technology and economics*. Cambridge: Cambridge University Press; 1982.
389. Rosenberg N. *Keynote address: challenges for the social sciences in the new millennium*. In *OECD Social Sciences and Innovation*. Paris: OECD; 2001.
390. Becker MH, Maiman LA. Strategies for enhancing patient compliance. *Journal Of Community Health*. 1980;6(2):113-135.
391. Nagy VT, Wolfe GR. Cognitive predictors of compliance in chronic disease patients. *Medical Care*. 1984;22(10):912-921.
392. Smith NA, Ley P, Seale JP, Shaw J. Health beliefs, satisfaction and compliance. *Patient Education And Counseling*. 1987;10(3):279-286.
393. Shikiar R, Rentz AM. Satisfaction with medication: an overview of conceptual, methodologic, and regulatory issues. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2004;7(2):204-215.
394. Murphy J, Coster G. Issues in patient compliance. *Drugs*. 1997;54(6):797-800.
395. O'Brien MK, Petrie K, Raeburn J. Adherence to medication regimens: updating a complex medical issue. *Medical Care Review*. 1992;49(4):435-454.
396. Rudd P. Clinicians and patients with hypertension: unsettled issues about compliance. *American Heart Journal*. 1995;130(3 Pt 1):572-579.
397. George J, Phun Y-T, Bailey MJ, Kong DCM, Stewart K. Development and validation of the medication regimen complexity index. *The Annals Of Pharmacotherapy*. 2004;38(9):1369-1376.
398. Hodgson TA, Meiners MR. Cost-of-illness methodology: a guide to current practices and procedures. *The Milbank Memorial Fund Quarterly Health And Society*. 1982;60(3):429-462.
399. Zhang W, Bansback N, Anis A. Measuring and valuing productivity loss due to poor health: A critical review. *Social Science and Medicine*. 2011;72(2):185-192.
400. Rothwell PM. Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clin Trials*. 2006;1(1):e9.
401. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ*. 2008;336(7649):876-880.
402. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical Research Ed)*. 2008;336(7650):924-926.

403. Grade Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454).
404. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical Research Ed)*. 2011;343:d5928-d5928.
405. Glasgow R, Vogt T, Boles S. Evaluating the Public Health Impact of Health Promotion Interventions: The RE-AIM Framework. *Am J Public Health Affairs*. 1999;89(9):1322-1327.
406. Bana e Costa CA, Vansnick J-C. MACBETH — An Interactive Path Towards the Construction of Cardinal Value Functions. *International Transactions in Operational Research*. 1994;1(4):489.
407. Bana e Costa C, De Corte J, Vansnick J. On the Mathematical Foundations of MACBETH. In: Greco S, Ehrgott M, Figueira J, eds. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer New York; 2016.
408. Bana e Costa CA, Vansnick J-C. Applications of the MACBETH Approach in the Framework of an Additive Aggregation Model. *Journal of Multi-Criteria Decision Analysis*. 1997;6(2):107-114.
409. Bana e Costa CA, Corrêa ÉC, De Corte J-M, Vansnick J-C. Facilitating bid evaluation in Public call for tenders: a socio-technical approach. *Omega*. 2002;30(3):227.
410. Bana e Costa CA, Oliveira MD. A multicriteria decision analysis model for faculty evaluation. *Omega*. 2012;40(4):424-436.
411. Bana e Costa C, Lourenço J, Oliveira M, Bana e Costa J. A Socio-technical Approach for Group Decision Support in Public Strategic Planning: The Pernambuco PPA Case. *Group Decision & Negotiation*. 2014;23(1):5-29.
412. Bana e Costa C, Lopes D, Oliveira M. Improving risk matrices using the MACBETH approach for multicriteria value measurement. In: *Proceedings of Maintenance Performance Measurement and Management (MPMM) Conference 2014*. Imprensa da Universidade de Coimbra; 2014.
413. Oliveira M, Bana e Costa C, Lopes D. Designing and exploring risk matrices with MACBETH. *International Journal of Information Technology & Decision Making*. 2015.
414. Peacock SJ, Richardson JRJ, Carter R, Edwards D. Priority setting in health care using multi- attribute utility theory and programme budgeting and marginal analysis (PBMA). *Social Science and Medicine*. 2007;64(4):897-910.
415. Gabrielli W, von Winterfeldt D. *Are importance weights sensitive to the range of alternatives in multiattribute utility measurement?* : University of Southern California, Social Science Research Institute;1978.
416. Phillips LD, Bana e Costa CA. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*. 2007;154(1):51-68.
417. Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, eds. *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht, The Netherlands: Springer; 2008.
418. de Bekker-Grob EW, Ryan M, Gerard K. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Health Economics*. 2012;21(2):145-172.

419. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2013;16(1):3-13.
420. Drummond M, Tarricone R, Torbica A. Assessing the added value of health technologies: reconciling different perspectives. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2013;16(1 Suppl):S7-S13.
421. Angelis A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: the Advance Value Framework. *Under Review*. 2016.
422. Franco LA, Montibeller G. Facilitated modelling in operational research. *European Journal of Operational Research*. 2010;205(3):489-500.
423. NICE. Bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. *Technology Appraisal Guidance 118*. 2007. <https://www.nice.org.uk/guidance/ta118>. Accessed 21 May 2015.
424. NICE. Cetuximab for the first-line treatment of metastatic colorectal cancer. 2009. <https://www.nice.org.uk/guidance/ta176>. Accessed 25 July 2016.
425. NICE. Bevacizumab in combination with oxaliplatin and either fluorouracil plus folinic acid or capecitabine for the treatment of metastatic colorectal cancer. *Technology Appraisal Guidance 212*. 2010. <https://www.nice.org.uk/guidance/ta118>. Accessed 21 May 2015.
426. NICE. Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first-line chemotherapy: Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy. *Technology Appraisal Guidance 242*. 2012. <https://www.nice.org.uk/guidance/ta242>. Accessed 21 May 2015.
427. NICE. Afibercept in combination with irinotecan and fluorouracil-based therapy for treating metastatic colorectal cancer that has progressed following prior oxaliplatin-based chemotherapy. *Technology Appraisal Guidance 307*. 2014. <https://www.nice.org.uk/guidance/ta307>. Accessed 21 May 2015.
428. NICE. Regorafenib for metastatic colorectal cancer after treatment for metastatic disease (terminated appraisal). *Technology Appraisal Guidance 334*. 2015. <https://www.nice.org.uk/guidance/ta334>. Accessed 21 May 2015.
429. Bana e Costa CA, Ensslin L, Corrêa EC, Vansnick J-C. Decision Support Systems in action: Integrated application in a multicriteria decision aid process. *European Journal of Operational Research*. 1999;113(2):315-335.
430. Saltz LB, Meropol NJ, Loehrer PJ, Sr., Needle MN, Kopit J, Mayer RJ. Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2004;22(7):1201-1208.
431. Ciardiello F, Tortora G. EGFR antagonists in cancer treatment. *The New England Journal Of Medicine*. 2008;358(11):1160-1174.
432. Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*. 2012;486(7404):532-536.

433. Van Cutsem E, Tabernero J, Lakomy R, et al. Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2012;30(28):3499-3506.
434. Price TJ, Peeters M, Kim TW, et al. Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority phase 3 study. *The Lancet Oncology*. 2014;15(6):569-579.
435. Hoyle M, Crathorne L, Peters J, et al. The clinical effectiveness and cost-effectiveness of cetuximab (mono- or combination chemotherapy), bevacizumab (combination with non-oxaliplatin chemotherapy) and panitumumab (monotherapy) for the treatment of metastatic colorectal cancer after first-line chemotherapy (review of technology appraisal No.150 and part review of technology appraisal No. 118): a systematic review and economic model. *Health Technology Assessment (Winchester, England)*. 2013;17(14):1-237.
436. EMA. Zaltrap (aflibercept) EPAR - Product Information. 2016. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002532/WC500139484.pdf. Accessed 25 July 2016.
437. EMA. Erbitux (cetuximab) EPAR - Product Information. 2014. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000558/WC500029119.pdf. Accessed 25 July 2016.
438. EMA. Vectibix (panitumumab) EPAR - Product Information. 2016. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000741/WC500047710.pdf. Accessed 25 July 2016.
439. World Health Organisation Collaborating Centre. ATC/DDD Index 2016. 2016; http://www.whocc.no/atc_ddd_index/. Accessed 25 July 2016.
440. ClinicalTrials.gov. US National Institutes of Health; 2016. <https://clinicaltrials.gov/>. Accessed 21 May 2015.
441. Wade R, Duarte A, Simmonds M, et al. *Aflibercept in combination with irinotecan and fluorouracil-based therapy for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy*. University of York;2013.
442. Hoyle M, Peters J, Crathorne L, et al. Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2013;16(2):288-296.
443. Phillips L. Decision Conferencing. In: Edwards W, Miles R, von Winterfeldt D, eds. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge: Cambridge University Press; 2007.
444. Phillips L. A theory of requisite decision models. *Acta Psychologica*. 1984;56:29-48.
445. Van Cutsem E, Peeters M, Siena S, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care

- alone in patients with chemotherapy-refractory metastatic colorectal cancer. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2007;25(13):1658-1664.
446. Schein E. *Process consultation revisited: building the helping relationship*. Reading: Addison–Wesley; 1999.
 447. Bana E Costa CA, De Corte J-M, Vansnick J-C. MACBETH. *International Journal of Information Technology & Decision Making*. 2012;11(2):359-387.
 448. Bana e Costa C, De Corte J, Vansnick J. M-MACBETH website. 2016; <http://www.m-macbeth.com/help/pdf/M-MACBETH%203.0.0%20Users%20Guide%20BETA.pdf>.
 449. Bana e Costa C, Vansnick J. The MACBETH Approach: Basic Ideas, Software, and an Application. In: Meskens N, Roubens M, eds. *Advances in Decision Analysis*. Springer Netherlands; 1999:131-157.
 450. Fasolo B, Bana e Costa CA. Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words. *Omega*. 2014;44:83-90.
 451. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Statistics In Medicine*. 2011;30(12):1419-1428.
 452. Ijzerman MJ, van Til JA, Snoek GJ. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. *The Patient*. 2008;1(4):265-272.
 453. Danner M, Hummel JM, Volz F, et al. Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. *International Journal Of Technology Assessment In Health Care*. 2011;27(4):369-375.
 454. Angelis A, Montibeller G, Hochhauser D, Kanavos P. Multiple Criteria Decision Analysis in the Context of Health Technology Assessment: a simulation exercise on metastatic colorectal cancer with multiple stakeholders in the English settings. *Under Review*. 2016.
 455. Huggins C, Hodges CV. Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *The Journal Of Urology*. 2002;167(2 Pt 2):948-951.
 456. Horwich A, Parker C, de Reijke T, Kataja V. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals Of Oncology: Official Journal Of The European Society For Medical Oncology / ESMO*. 2013;24 Suppl 6:vi106-vi114.
 457. Hussain M, Tangen CM, Berry DL, et al. Intermittent versus continuous androgen deprivation in prostate cancer. *The New England Journal Of Medicine*. 2013;368(14):1314-1325.
 458. Samson DJ, Seidenfeld J, Schmitt B, et al. Systematic review and meta-analysis of monotherapy compared with combined androgen blockade for patients with advanced prostate carcinoma. *Cancer*. 2002;95(2):361-376.
 459. Tannock IF, de Wit R, Berry WR, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *The New England Journal Of Medicine*. 2004;351(15):1502-1512.
 460. Berthold DR, Pond GR, Soban F, de Wit R, Eisenberger M, Tannock IF. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced

- prostate cancer: updated survival in the TAX 327 study. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2008;26(2):242-245.
461. de Bono JS, Oudard S, Ozguroglu M, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet*. 2010;376(9747):1147-1154.
 462. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. *The New England Journal Of Medicine*. 2011;364(21):1995-2005.
 463. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *The New England Journal Of Medicine*. 2012;367(13):1187-1197.
 464. Bianchini D, Lorente D, Rodriguez-Vida A, et al. Antitumour activity of enzalutamide (MDV3100) in patients with metastatic castration-resistant prostate cancer (CRPC) pre-treated with docetaxel and abiraterone. *European Journal Of Cancer (Oxford, England: 1990)*. 2014;50(1):78-84.
 465. Loriot Y, Bianchini D, Ileana E, et al. Antitumour activity of abiraterone acetate against metastatic castration-resistant prostate cancer progressing after docetaxel and enzalutamide (MDV3100). *Annals Of Oncology: Official Journal Of The European Society For Medical Oncology / ESMO*. 2013;24(7):1807-1812.
 466. NICE. Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen. *Technology Appraisal Guidance 259*. 2012. <https://www.nice.org.uk/guidance/ta259>. Accessed 21 May 2015.
 467. NICE. Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel - containing regimen. *Technology Appraisal Guidance 316*. 2014. <https://www.nice.org.uk/guidance/ta316>. Accessed 21 May 2015.
 468. Ryan CJ, Smith MR, de Bono JS, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. *The New England Journal Of Medicine*. 2013;368(2):138-148.
 469. Beer TM, Armstrong AJ, Rathkopf DE, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. *The New England Journal Of Medicine*. 2014;371(5):424-433.
 470. Sweeney S, Chen Y, Carducci M, et al. Impact on overall survival (OS) with chemohormonal therapy versus hormonal therapy for hormone-sensitive newly metastatic prostate cancer (mPrCa): An ECOG-led phase III randomized trial. *J Clin Oncol*. 2014;32(5s):suppl; abstr LBA2.
 471. James N, Sydes M, Mason M, et al. Docetaxel and/or zoledronic acid for hormone-naïve prostate cancer: First overall survival results from STAMPEDE (NCT00268476). *J Clin Oncol*. 2015;33(suppl):abstr 5001.
 472. NICE. Cabazitaxel for hormone-refractory metastatic prostate cancer previously treated with a docetaxel-containing regimen. *Technology Appraisal Guidance 255*. 2012. <https://www.nice.org.uk/guidance/ta255>. Accessed 21 May 2015.
 473. TLV. Designation 2775/2013. 2014. http://www.tlv.se/Upload/Beslut_2014/bes140527-xtandi.pdf. Accessed 21 May 2015.

474. TLV. Designation 4774/2014. 2015.
http://www.tlv.se/upload/beslut_2015/bes150521-zytiga.pdf. Accessed 21 May 2015.
475. Bana e Costa C, Beinat E. *Model-structuring in public decision-aiding*. London: London School of Economics and Political Science;2005.
476. Fizazi K, Scher HI, Molina A, et al. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *The Lancet Oncology*. 2012;13(10):983-992.
477. EMA. Zytiga (abiraterone) EPAR - Product Information. 2016.
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002321/WC500112858.pdf. Accessed 26 July 2016.
478. EMA. Jevtana (cabazitaxel) EPAR - Product Information. 2016.
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002018/WC500104764.pdf. Accessed 26 July 2016.
479. EMA. Xtandi (enzalutamide) EPAR - Product Information. 2016.
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002639/WC500144996.pdf. Accessed 26 July 2016.
480. Connock M, Cummins E, Shyangdan D, Hall B, Grove A, Clarke A. *Abiraterone acetate for the treatment of metastatic, castrate-resistant prostate cancer following previous cytotoxic chemotherapy: A Single Technology Appraisal*. Warwick Evidence;2011.
481. Riensa R, Ramaekers B, Tomini F, et al. *Abiraterone for the treatment of chemotherapy naïve metastatic castration-resistant prostate cancer: a Single Technology Appraisal*. Kleijnen Systematic Reviews Ltd;2013.
482. Collins R, Fenwick E, Trowman R, et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. *Health Technology Assessment (Winchester, England)*. 2007;11(2):iii.
483. Kearns B, Lloyd Jones M, Stevenson M, Littlewood C. Cabazitaxel for the Second-Line Treatment of Metastatic Hormone-Refractory Prostate Cancer: A NICE Single Technology Appraisal. *Pharmacoeconomics*. 2013;31(6):479-488.
484. Burström K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the EQ-5D. *Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation*. 2001;10(7):621-635.
485. Karlsberg Schaffer S, Sussex J, Devlin N, Walker A. Local health care expenditure plans and their opportunity costs. *Health Policy*. 2015;119(9):1237-1244.
486. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and care excellence cost- effectiveness threshold. *Health Technology Assessment*. 2015;19(14):1-503.
487. Hummel MJM, Volz F, van Manen JG, et al. Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment. *The Patient*. 2012;5(4):225-237.

488. Wilson E, Sussex J, Macleod C, Fordham R. Prioritizing health technologies in a Primary Care Trust. *Journal Of Health Services Research & Policy*. 2007;12(2):80-85.
489. Freeman R. *Strategic management: a stakeholder approach*. New York: Cambridge University Press; 2010.
490. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*. 2011;14(4):417-428.
491. Naci H, Brugts JJ, Fleurence R, Tsoi B, Toor H, Ades AE. Comparative benefits of statins in the primary and secondary prevention of major coronary events and all-cause mortality: a network meta-analysis of placebo-controlled and active-comparator trials. *European Journal Of Preventive Cardiology*. 2013;20(4):641-657.
492. Naci H, Brugts JJ, Fleurence R, Ades AE. Comparative effects of statins on major cerebrovascular events: a multiple-treatments meta-analysis of placebo-controlled and active-comparator trials. *QJM: Monthly Journal Of The Association Of Physicians*. 2013;106(4):299-306.
493. Naci H, Brugts J, Ades T. Comparative tolerability and harms of individual statins: a study-level network meta-analysis of 246 955 participants from 135 randomized, controlled trials. *Circulation Cardiovascular Quality And Outcomes*. 2013;6(4):390-399.
494. van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multicriteria benefit-risk assessment using network meta-analysis. *Journal Of Clinical Epidemiology*. 2012;65(4):394-403.
495. Kirkwood C. *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Belmont, California: Duxbury Press; 1997.
496. Bana e Costa CA, Carnero MC, Oliveira MD. A multi-criteria model for auditing a Predictive Maintenance Programme. *European Journal of Operational Research*. 2012;217(2):381-393.
497. Dyer J. MAUT – Multiattribute utility theory. In: Figueira J, Greco S, Ehrgott M, eds. *Multiple Criteria Decision Analysis*. Springer; 2005:265-296.
498. Montibeller G, Winterfeldt D. Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis: An International Journal*. 2015;35(7):1230-1251.
499. Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Applied Health Economics And Health Policy*. 2011;9(5):331-347.
500. EMA. Adaptive pathways: key learnings and next steps. 2016; http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2016/08/news_detail_002586.jsp&mid=WC0b01ac058004d5c1. Accessed 10 September 2016.
501. EMA. Cooperation between regulators and HTA bodies creates synergies. 2016; http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2016/04/news_detail_002510.jsp&mid=WC0b01ac058004d5c1. Accessed 10 September 2016.

502. Fried TR. Shared Decision Making--Finding the Sweet Spot. *The New England Journal Of Medicine*. 2016;374(2):104-106.
503. Naci H. Communication of Treatment Rankings Obtained From Network Meta-Analysis Using Data Visualization. *Circ Cardiovasc Qual Outcomes*. 2016.
504. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Medical Research Methodology*. 2003;3:28-28.
505. Kass MA, Meltzer DW, Gordon M, Cooper D, Goldberg J. Compliance with topical pilocarpine treatment. *American Journal Of Ophthalmology*. 1986;101(5):515-523.
506. Gordon M, Kass M. Validity of standard compliance measures in glaucoma compared with an electronic eyedrop monitor. In: Cramer J, Spilker B, eds. *Patient compliance in medical practice and clinical trials*. New York: Raven Press; 1991:163-173.
507. Cramer J, Mattson R. Monitoring compliance with antiepileptic drug therapy. In: Cramer J, Spilker B, eds. New York: Raven Press; 1991:123-137.
508. Kruse W, Weber E. Dynamics of drug regimen compliance--its assessment by microprocessor-based monitoring. *European Journal Of Clinical Pharmacology*. 1990;38(6):561-565.
509. Jonker DJ, O'Callaghan CJ, Karapetis CS, et al. Cetuximab for the treatment of colorectal cancer. *The New England Journal Of Medicine*. 2007;357(20):2040-2048.
510. Giantonio BJ, Catalano PJ, Meropol NJ, et al. Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200. *Journal Of Clinical Oncology: Official Journal Of The American Society Of Clinical Oncology*. 2007;25(12):1539-1544.

THESIS APPENDICES

Appendix to Chapter 2

Main types of Economic Evaluation:

	CBA	CEA	CUA
Objective	allocative efficiency (economic welfare theory)	technical efficiency	technical or allocative efficiency (extra-welfarism theory)
Outcomes	single or multiple and not necessarily common across alternatives in monetary units	single type of effect that is common between the alternatives being compared (e.g. 'life years gained')	utility associated with the effects of comparators being assessed (e.g. QALY), which can be single or multiple and not necessarily common across the alternatives
Costs	monetary units	monetary units	monetary units

Perspective	societal	Decision-maker	health care policy maker
Result	net benefit ratio (ratio of cost to benefits)	cost per unit of outcome (e.g. cost per life year gained), or effect per unit of cost (e.g. life year gained per dollar spent)	cost per utility-adjusted health outcome (e.g. QALY)

Sources: ^{108,118,121}

Appendix to Chapter 5

The available clinical evidence used for the assessment of therapeutic benefit will in most cases be sourced from explanatory trials that reflect ideal conditions, conducted on a highly selected group of patients and under strictly controlled environments while ensuring regimen compliance. In contrast, pragmatic trials are conducted under real world settings and on patients representing the full population spectrum which may show varying compliance ⁵⁰⁴. Indeed, evidence from electronic monitoring for a range of diseases including hypertension, glaucoma, seizure disorders and others, indicate that good adherence to prescribed regimens is only observed in between 50% and 60% of patients, with 5% to 10% adhering poorly and 30% to 45% adhering to an intermediate but widely variable degree ^{396,505-508}. Ergo, and unless the clinical evidence are coming from pragmatic trials resembling real world conditions, acknowledging that there is an impact on health outcomes from (un)ease and/or (un)convenient regimens could act as an adjustment or “fixture” towards the reflection of a more realistic clinical picture.

Appendix to Chapter 6

6.1 Methodological Framework

At first, as part of the problem structuring phase, the decision problem and the aims of the analysis are defined, and the relevant decision-makers and other key stakeholders are identified. Next, as part of the model-building phase, objectives and/or relevant criteria are identified in order to reflect decision-makers' goals and areas of concern, and attributes are selected to operationalise the criteria. In addition, under the same phase, selection of the alternative options takes place and evidence on their performance across the selected criteria is identified. Following that, under the model assessment phase, the performance of options against the criteria is assessed (i.e. scoring) and criteria are weighted according to their relative importance (i.e. weighing). Subsequently, as part of the appraisal phase, scores and weights are combined in order to produce overall WPV scores, taking the form of a value index (i.e. aggregation). In combination with sensitivity analysis, the results are examined and their robustness is determined. Finally, as part of action planning, the outcome of the analysis can be used to inform resource allocation decisions, of a coverage or pricing nature.

6.2 Evidence Considered and Alternative Treatments Compared

As part of TA242, for the case of cetuximab and panitumumab, NICE considered clinical evidence coming from two open label, Phase 3 RCTs respectively; the first one investigating the use of cetuximab plus best supportive care (BSC) compared to BSC alone¹²⁵ (*CO.17* trial)⁵⁰⁹ and the second one investigating the use of panitumumab plus BSC compared to BSC alone (*AMGEN* trial)⁴⁴⁵, in patients with chemotherapy-refractory mCRC. For the case of bevacizumab, as part of TA242, only one RCT had been identified investigating bevacizumab as a second line treatment (*E3200* trial)⁵¹⁰. However in that trial, bevacizumab was administered in combination with an oxaliplatin-containing chemotherapy which was outside the appraisal's scope, and hence outside the scope of our analysis.

As part of TA307, the clinical evidence for aflibercept was taken from a prospective multinational, randomized, double-blind, parallel-arm, phase 3 study investigating the addition of aflibercept to FOLFIRI in patients with mCRC previously treated with an oxaliplatin-based regimen (*VELOUR* trial)⁴³³.

Finally, for the case of regorafenib, no clinical evidence was considered as part of TA334 because no evidence submission was received from the manufacturer and the appraisal was terminated early.

No indirect comparison was conducted given the lack of a common comparator among the three treatments of interest among the above clinical studies. A mixed-treatment comparison lied outside the aim of the simulation exercise which was to operationalise the value framework in place through the elicitation of preferences across a range of explicit criteria from a group of stakeholders. As a result, for the case of cetuximab and panitumumab clinical evidence was used from a latest head to head, open label, randomised, multicentre Phase 3 non-inferiority study directly comparing both treatments (*ASPECCT* trial)⁴³⁴, whereas for the case of aflibercept in combination with non-oxaliplatin based chemotherapy evidence was used from the same clinical study that NICE considered. However, data from the BSC comparator arms of the two trials that NICE considered as part of TA242 for the case of cetuximab and panitumumab (*CO.17*, *AMGEN*)^{445,509}, were used for the purpose of setting the reference levels on the attributes.

¹²⁵ No evidence were submitted to NICE for cetuximab in combination with chemotherapy, therefore this combination fell outside the scope of our exercise too.

6.3 MCDA Technique

MACBETH uses seven semantic categories ranging between “no difference” to “extreme difference”, in order to distinguish between the value of different attribute levels. Based on these qualitative judgements of difference and, by analysing judgmental inconsistencies, it facilitates the move from ordinal preference modeling, a cognitively less demanding elicitation of preferences, to a quantitative value function. An example of the type of questioning being asked would be “What do you judge is the difference of value between x' and x'' ?” where x' and x'' are two different attribute levels of attribute x , across the plausible range (i.e. $x_* \leq x', x'' \leq x^*$). The approach has evolved through the course of theoretical research and real world practical applications, making it an interactive decision support system that facilitates decision-makers’ communication.

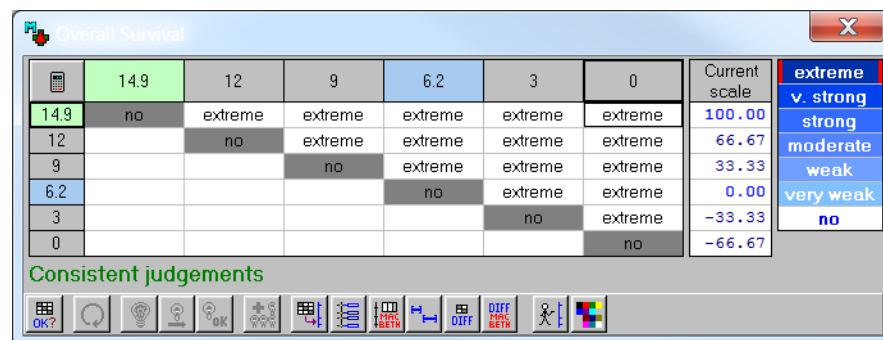
Following the elicitation of value functions, criteria baseline weights can be elicited. Questions of direct importance for a criterion such as “*How important is a given criterion?*” are known to be as one of the most common mistakes when making value trade-offs because they are assessing them independent of the respective attribute ranges ²⁰². In contrast, indirect weighting technique that assess value trade-offs in tandem with the respective ranges of attributes should be employed. For example, the quantitative swing weighting technique asks for judgments of relative value between ‘swings’ (i.e. changes) from standard lower level x_* to higher reference level x^* on each x -th attribute) taking the form “*How would you rank the relative importance of the criteria, considering their attributes ranges relative to 100 for the highest-ranked criterion considering its range?*”. Each swing, i.e. a relative change from a lower attribute level to a higher attribute level, is valued between 0 and 100, with the most valuable swing anchored as 100 ⁵⁴. Normalised weights are then calculated, as a proportion of each swing weight, so the normalised weights summed to 100%. Instead, relative attribute weights were calculated using an alternative qualitative swing weighting protocol, by using the MACBETH procedure to elicit the differences in attractiveness between the lower and higher reference levels of the different attributes, initially at individual level and then at criteria cluster level (i.e. by considering multiple attribute swings on the same time) ^{407,447}.

Finally criteria preference value scores and the respective weights can be combined together through an additive aggregation approach as described in Eq. X (if

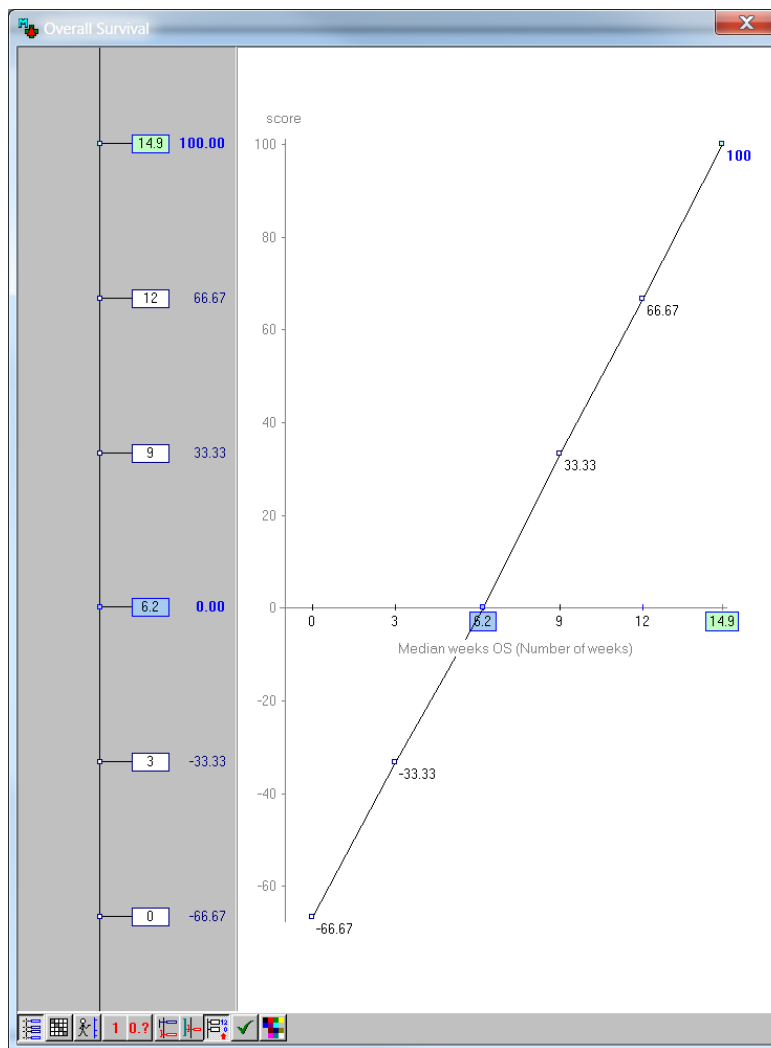
the adequate conditions of complete and transitive preferences are met as well as multi-attribute preferential independence conditions – see ⁵⁴.

6.4 Options Performance, Criteria Weights and Overall Value Rankings¹²⁶

6.4.1 Example of a value judgement matrix for the Overall Survival attribute and its conversion into a value function



¹²⁶ images produced using the M-MACBETH (beta) software version 3.0.0



Caption: In this example, the question asked for the case of OS was the following: “What do you judge to be the difference of value between 0 months OS and 14.9 months OS? No difference, very weak, weak, moderate, strong, very strong, or extreme?” Once a consensus was reached, the next question came along: “What do you judge to be the difference of value between 3 months OS and 14.9 months OS? No difference, very weak, weak, moderate, strong, very strong, or extreme?” The same process was followed until value judgments for all the different combinations of attribute levels were elicited, filling in the different rows from the right-hand side (i.e. lower range) to the left-hand side (i.e. higher range), top to bottom.

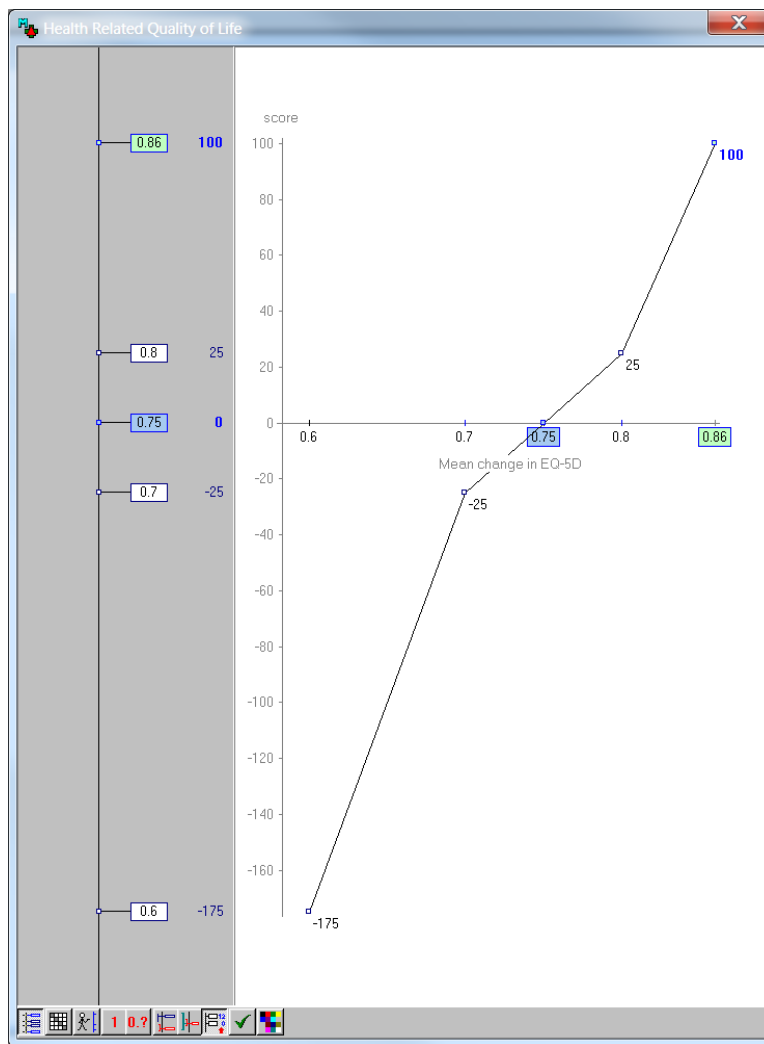
6.4.2 Example of a value judgement matrix for the Health Related Quality of Life

attribute and its conversion into a value function

The figure is a window titled "Health Related Quality of Life" showing a value judgement matrix. The matrix compares attribute levels 0.86, 0.8, 0.75, 0.7, and 0.6. The current scale is shown on the right, ranging from -175 to 100. The matrix is as follows:

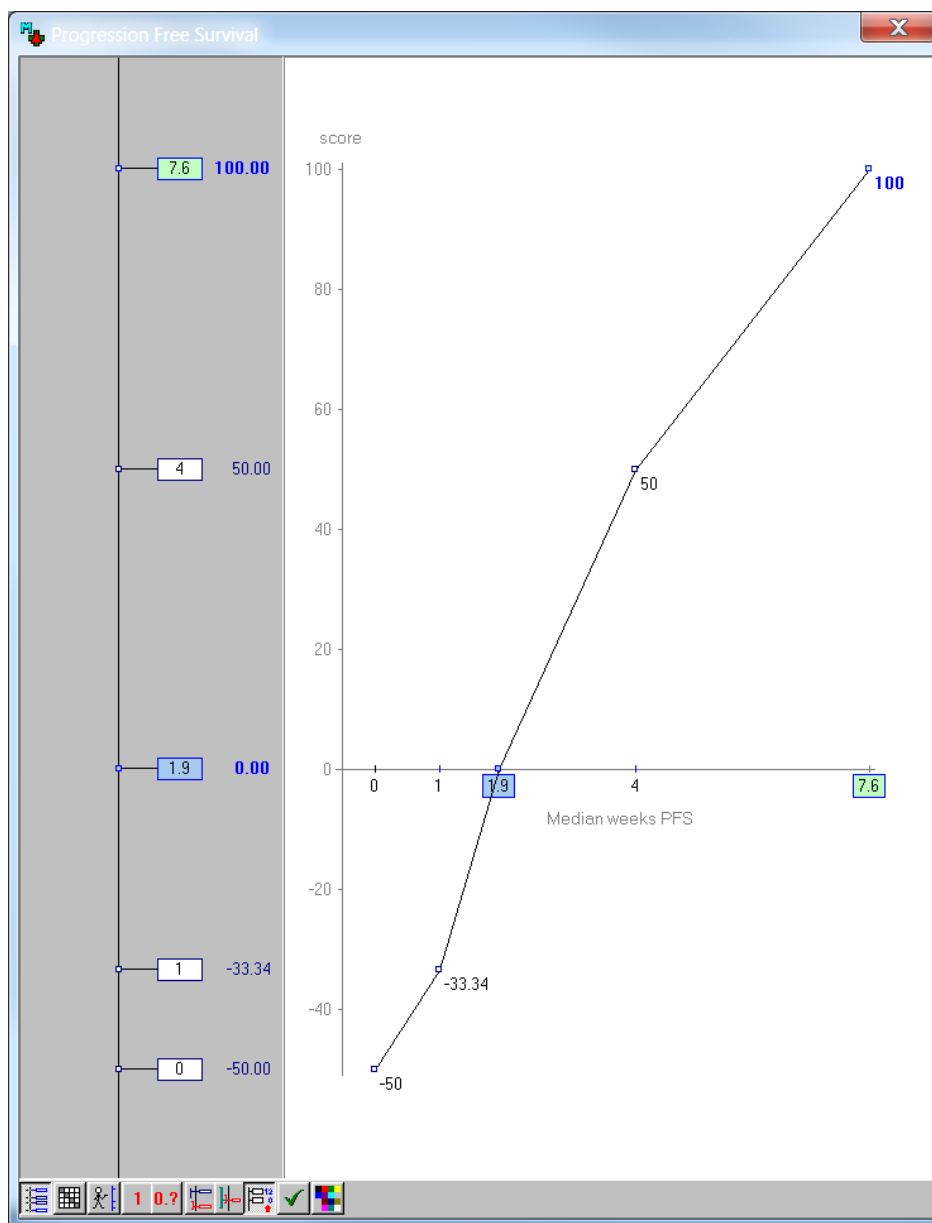
	0.86	0.8	0.75	0.7	0.6	Current scale	
0.86		no	mod-strg	strg-vstr	strg-vstr	100	extreme
0.8			no	positive	positive	25	v. strong
0.75				no	positive	0	strong
0.7					no	-25	moderate
0.6						-175	weak
							very weak
							no

Consistent judgements

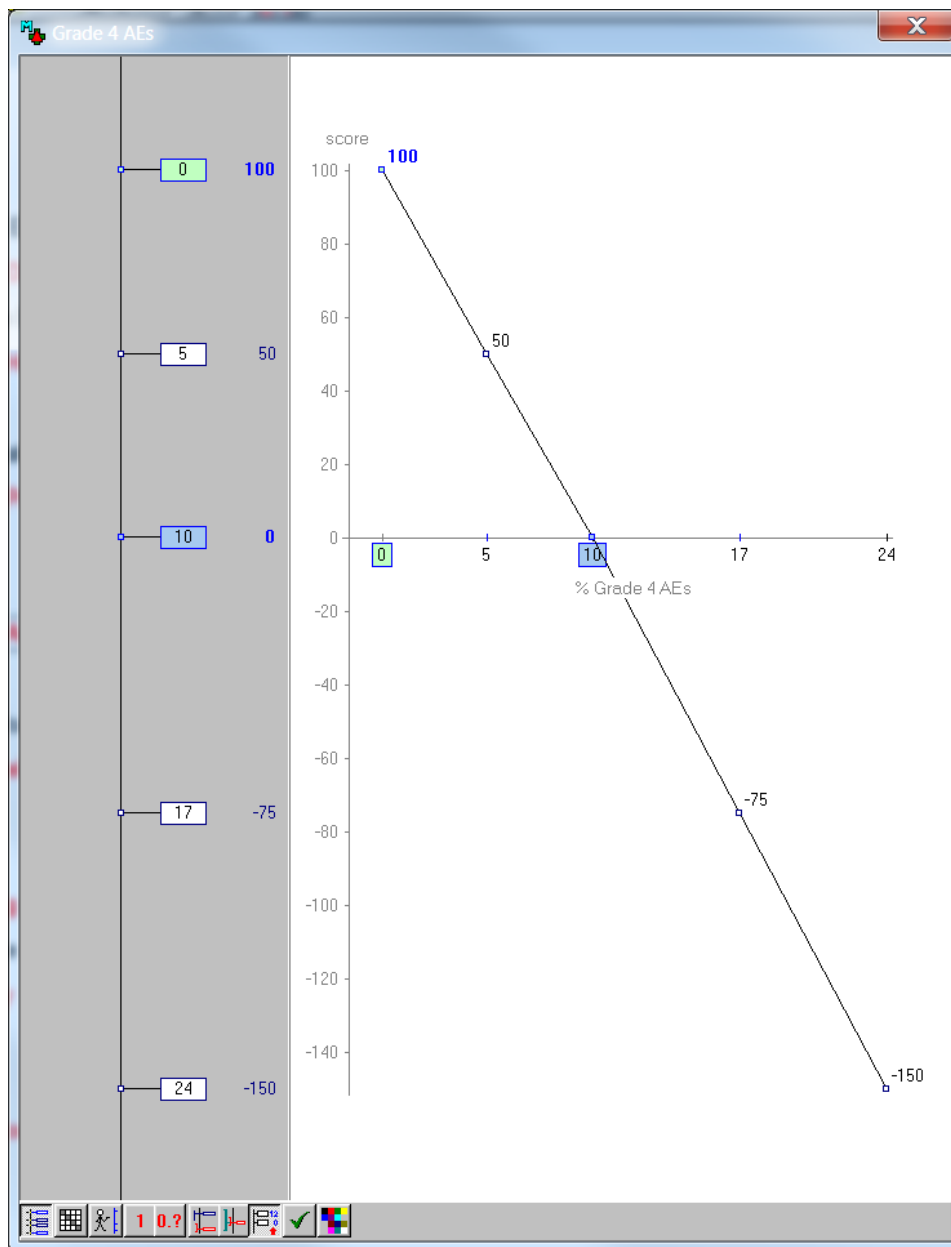


Caption: In this example, the question asked for the case of HRQoL was the following: “What do you judge to be the difference of value between 0.6 EQ-5D index score OS and 0.86 EQ-5D index score? No difference, very weak, weak, moderate, strong, very strong, or extreme?” Once a consensus was reached, the next question came along: “What do you judge to be the difference of value between 0.7 EQ-5D index score OS and 0.86 EQ-5D index score? No difference, very weak, weak, moderate, strong, very strong, or extreme?” The same process was followed until value judgments for all the different combinations of attribute levels were elicited, filling in the different rows from the right-hand side (i.e. lower range) to the left-hand side (i.e. higher range), top to bottom.

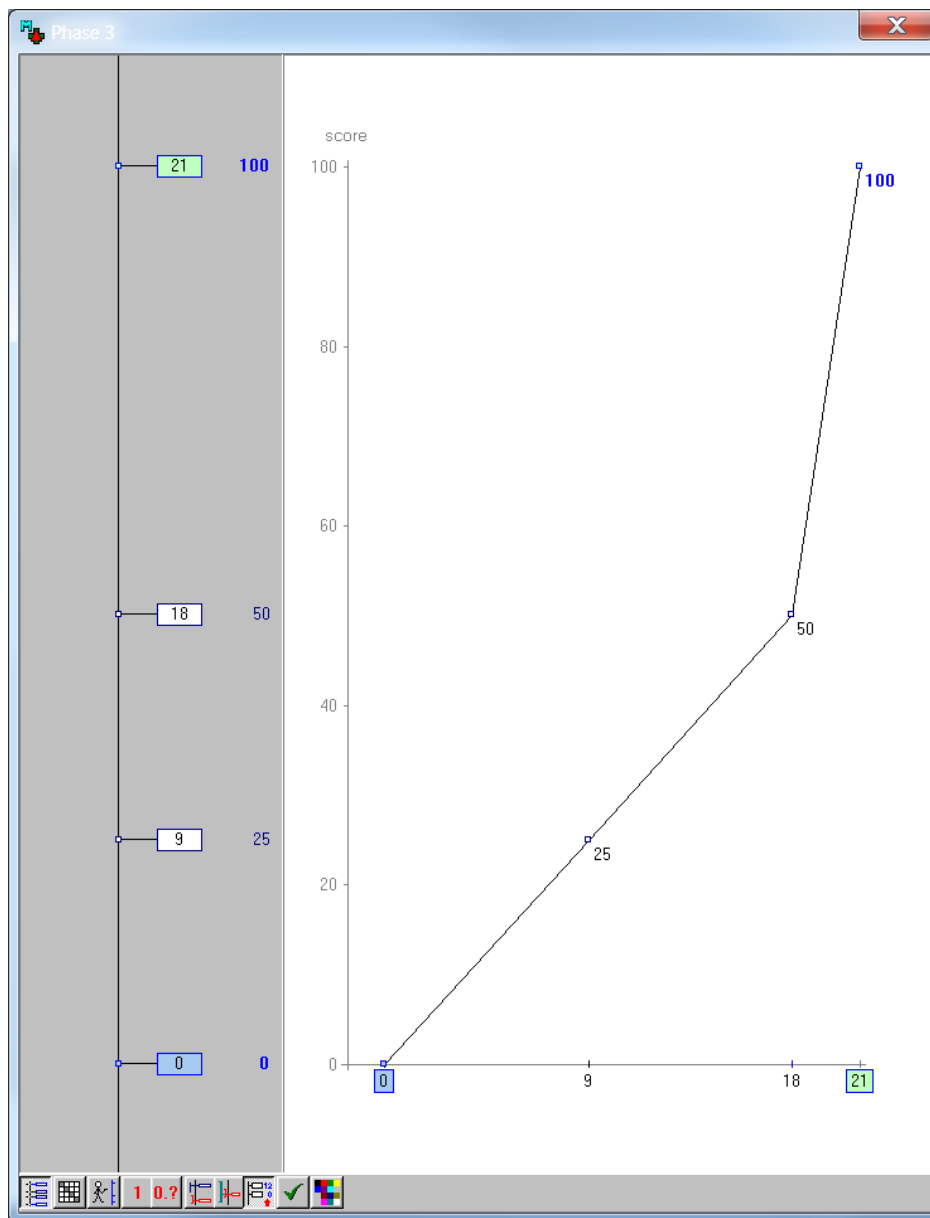
6.4.3 Value scale for the Progression Free Survival attribute



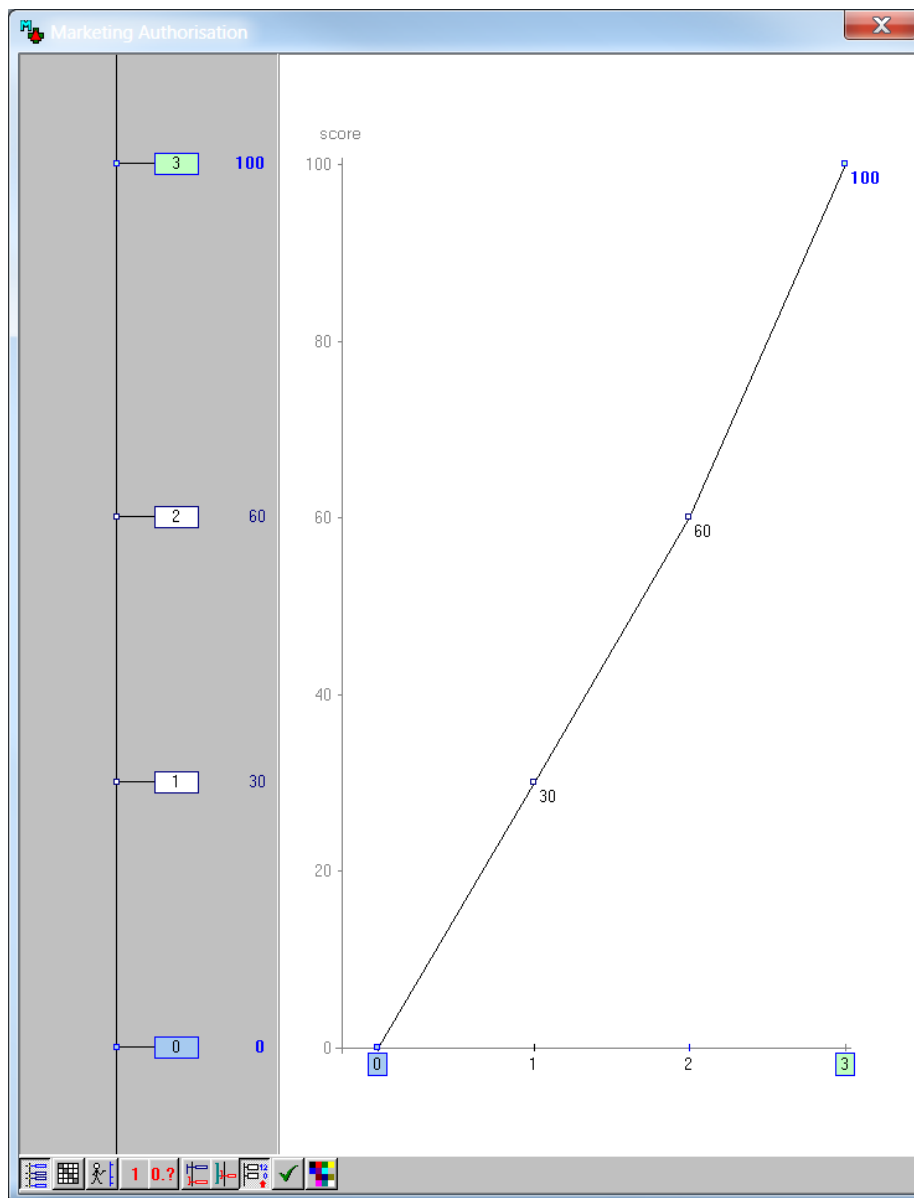
6.4.4 Value scale for Grade 4 AEs attribute



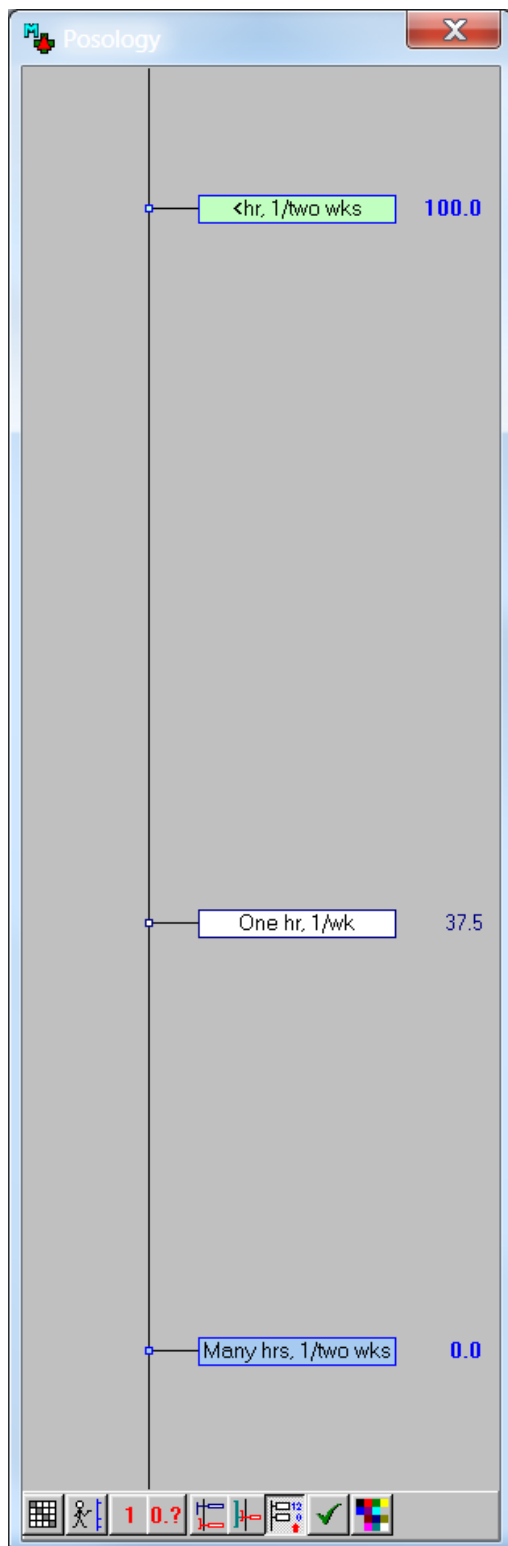
6.4.5 Value scale for ATC L4 attribute



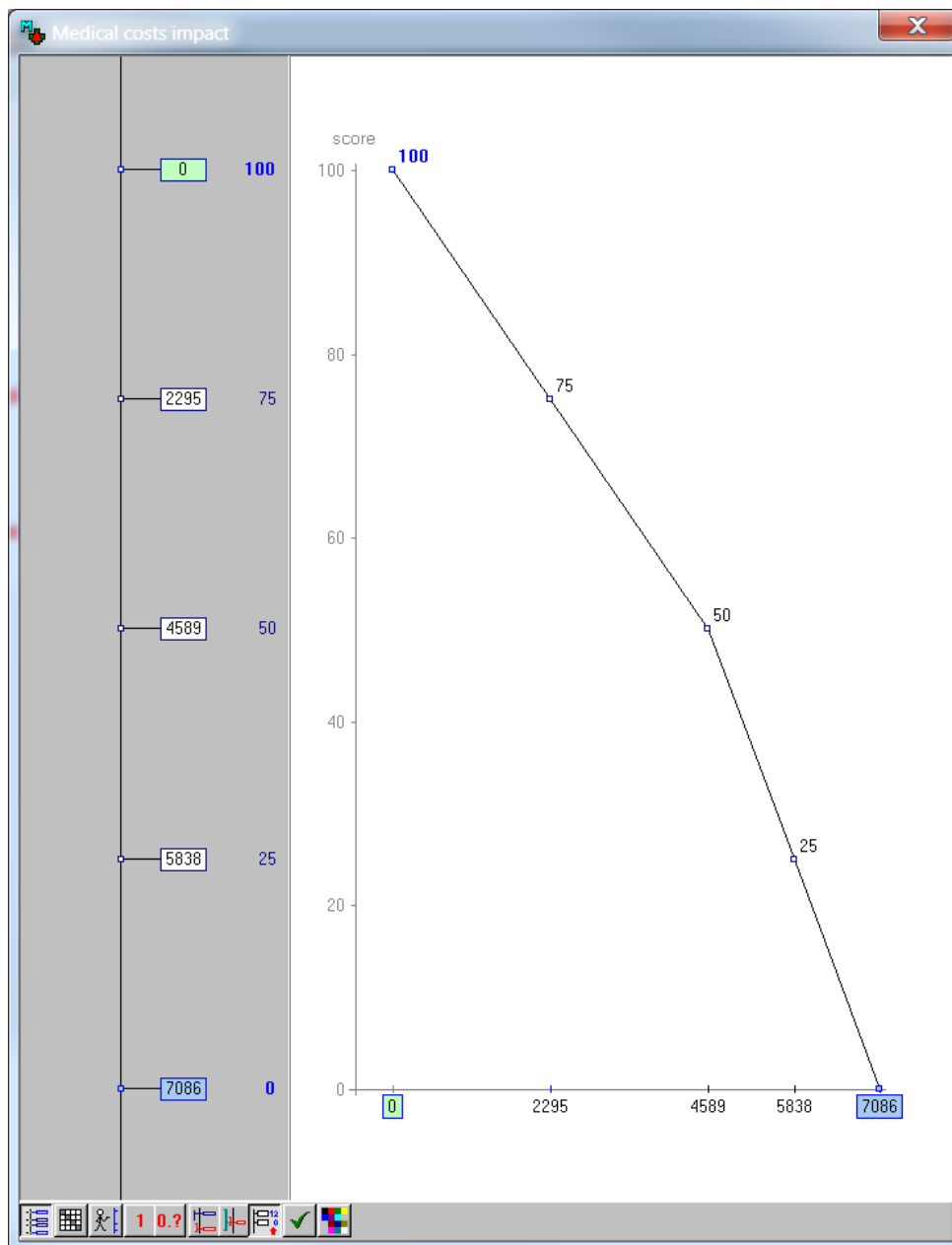
6.4.7 Value scale for Marketing Authorisation attribute



6.4.8 Value scale for Posology attribute

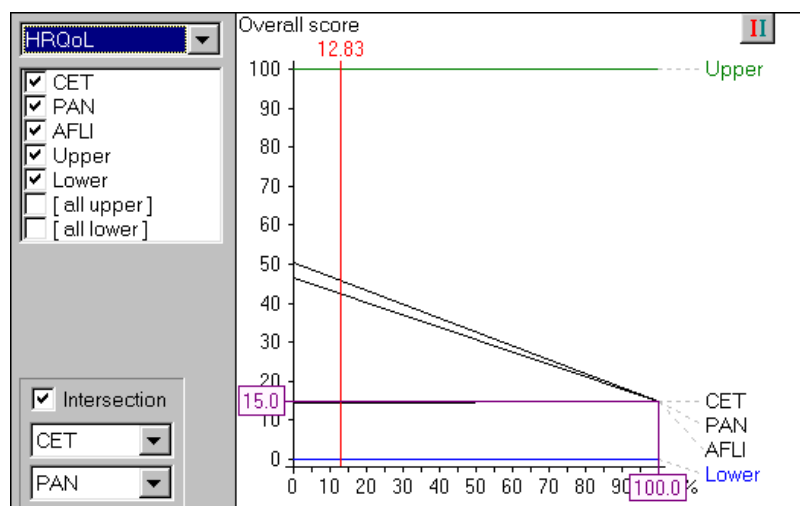
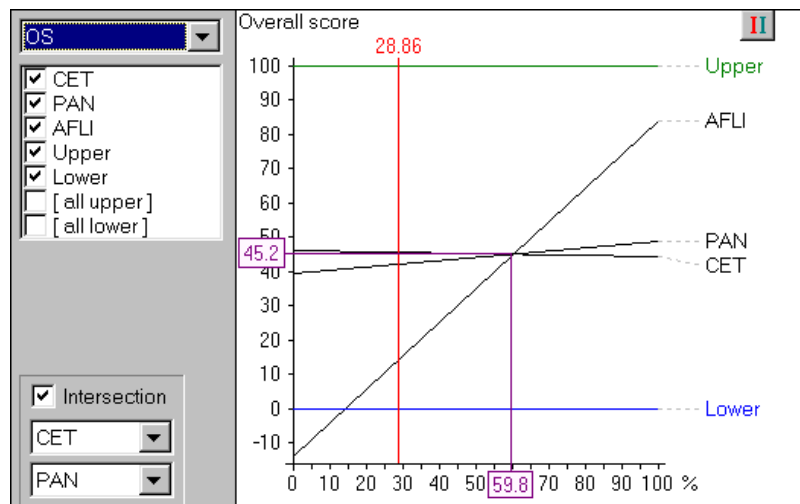


6.4.9 Value scale for Medical costs impact attribute

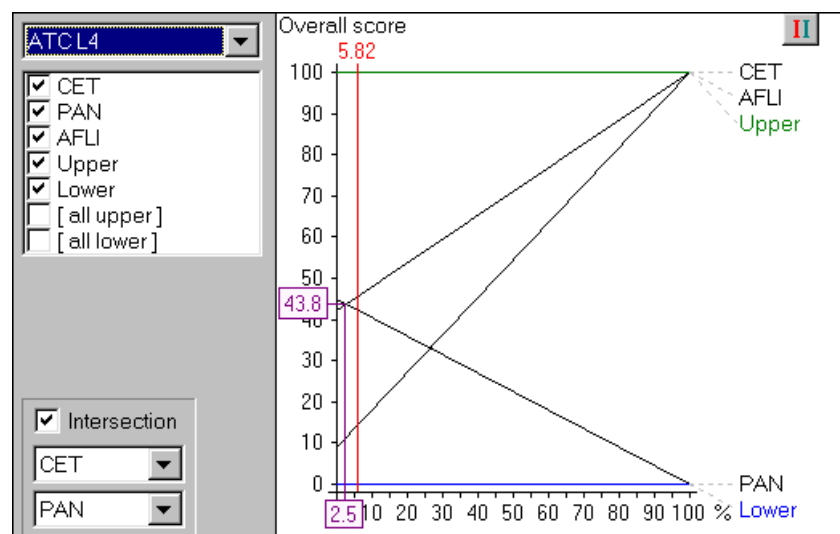
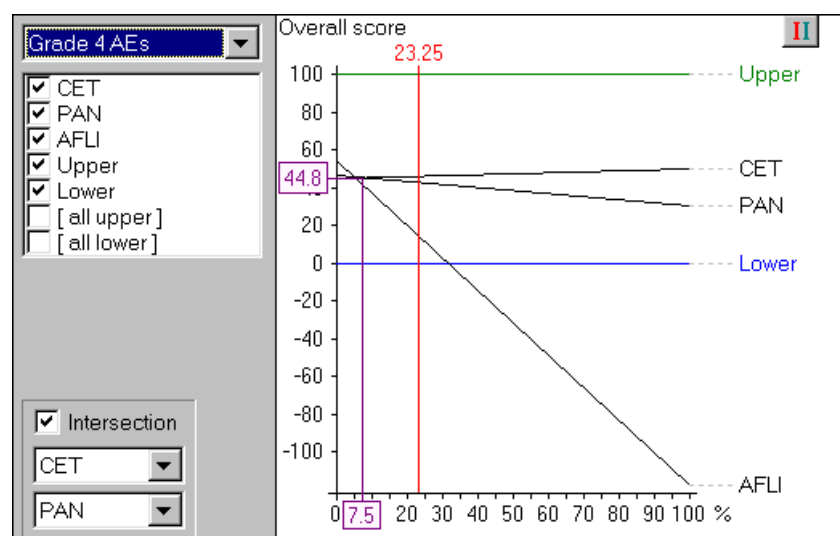
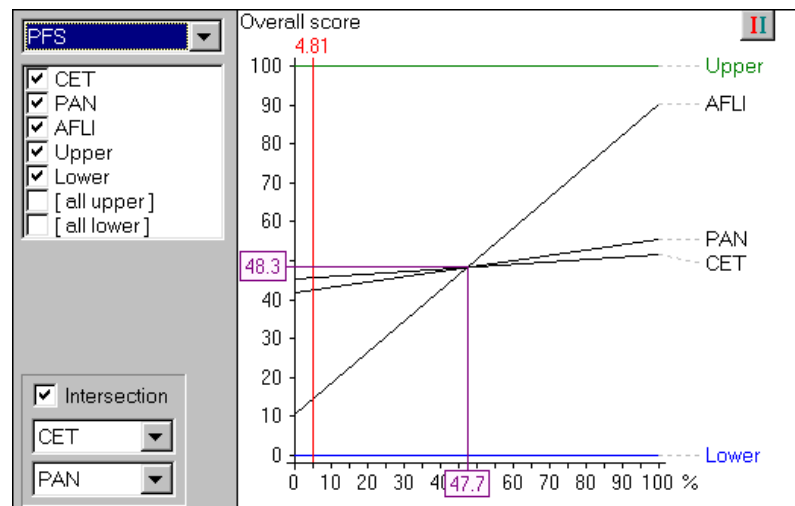


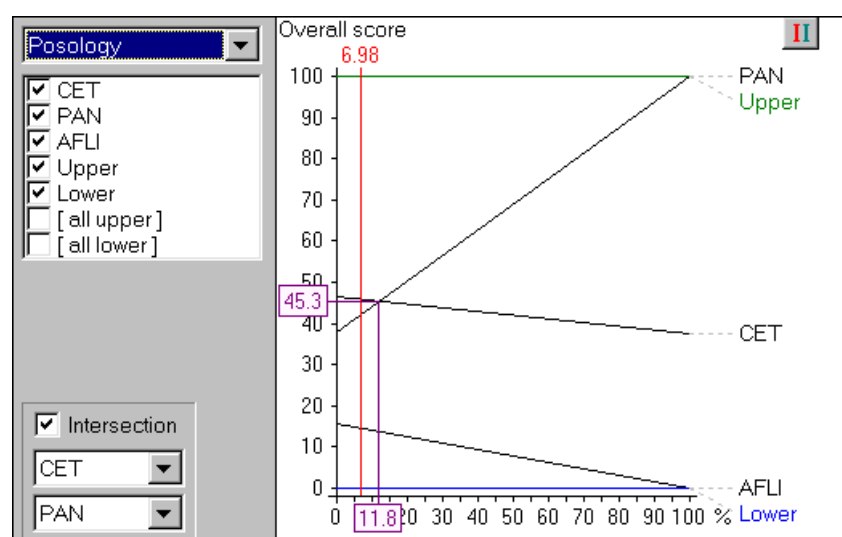
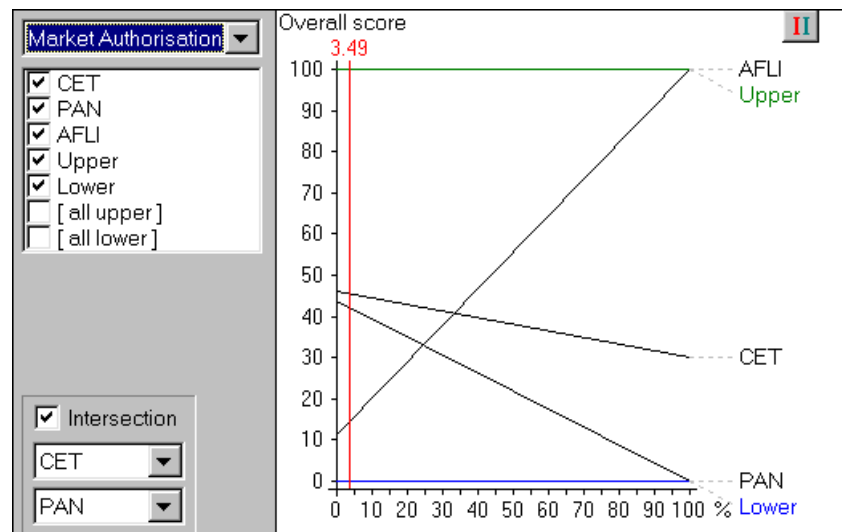
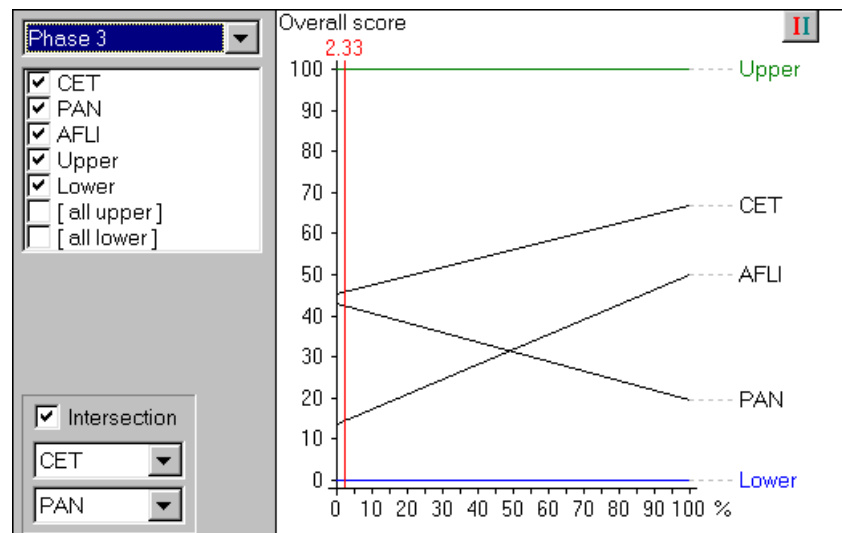
6.5 Sensitivity and Robustness Analysis¹²⁷

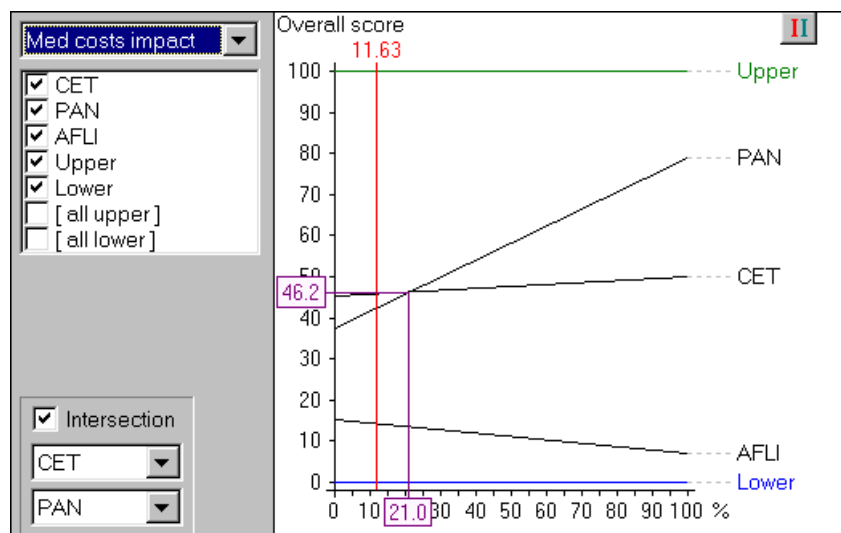
6.5.1 Sensitivity analysis on weights for Cetuximab versus Panitumumab



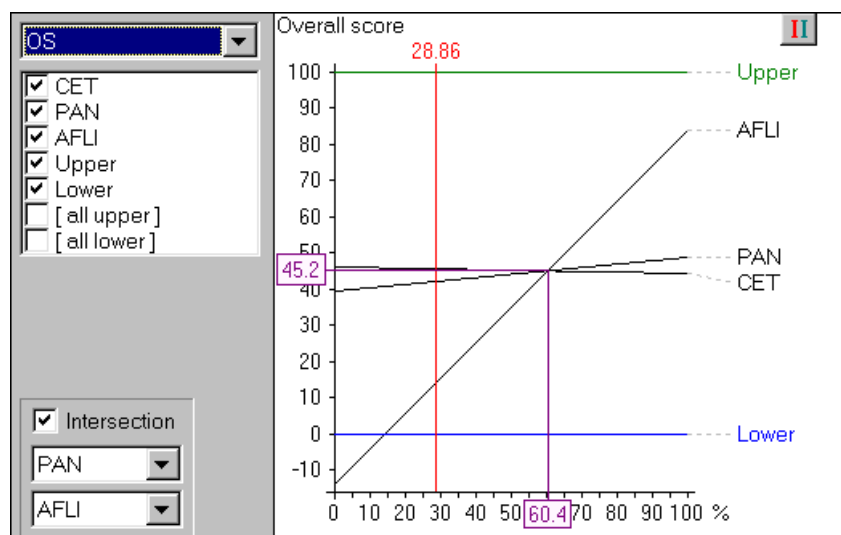
¹²⁷ images produced using the M-MACBETH (beta) software version 3.0.0

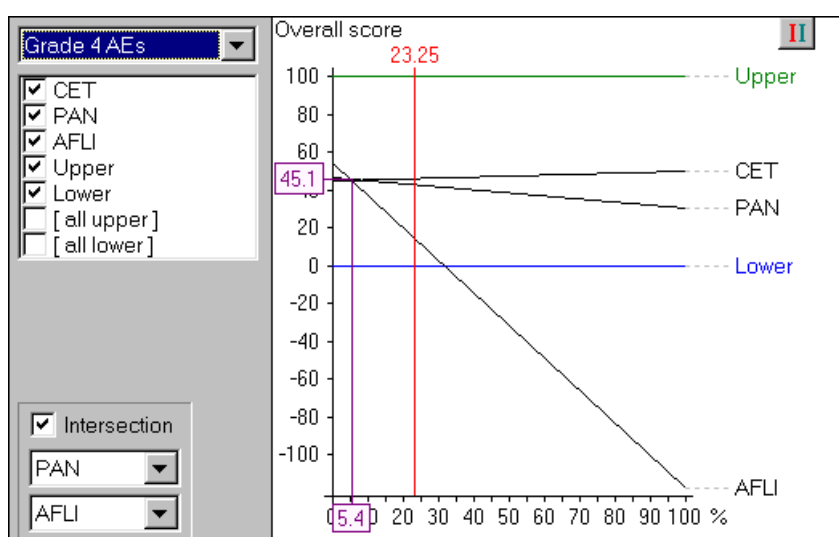
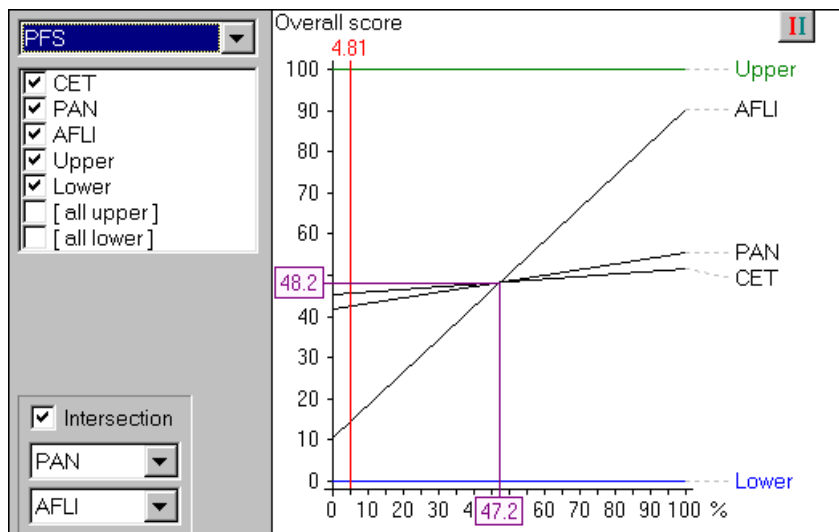
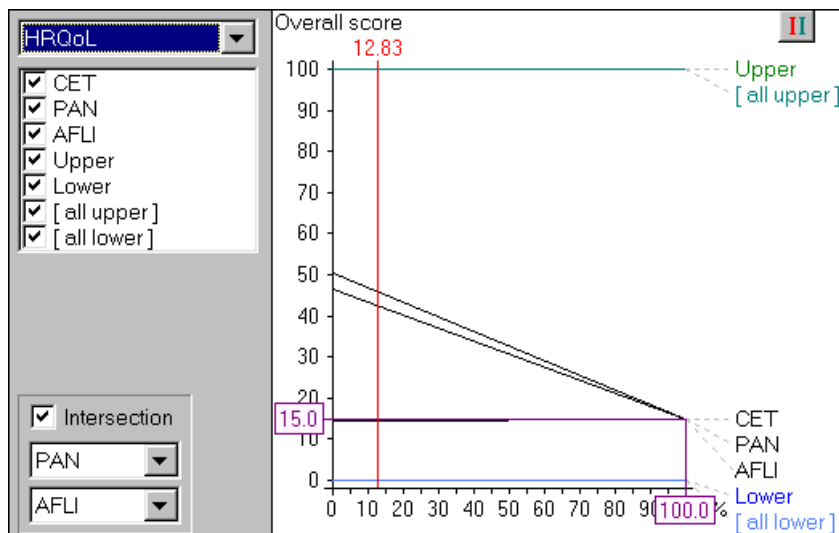


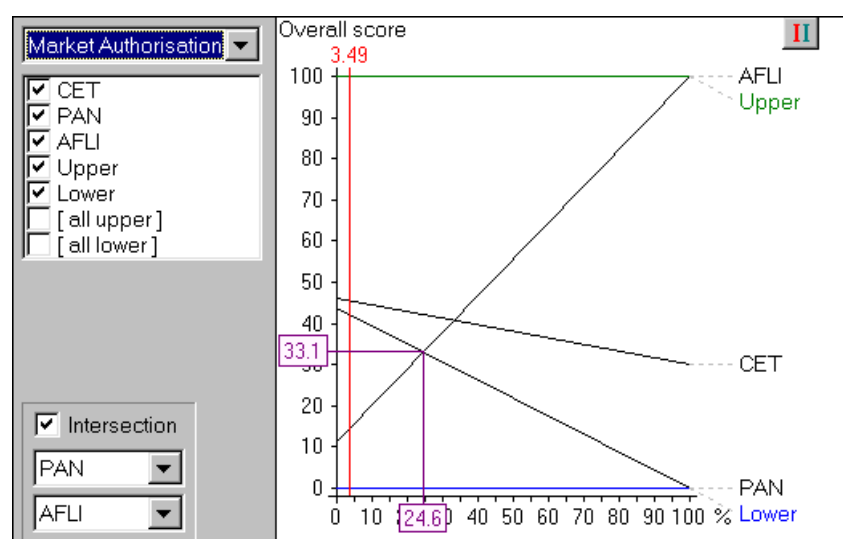
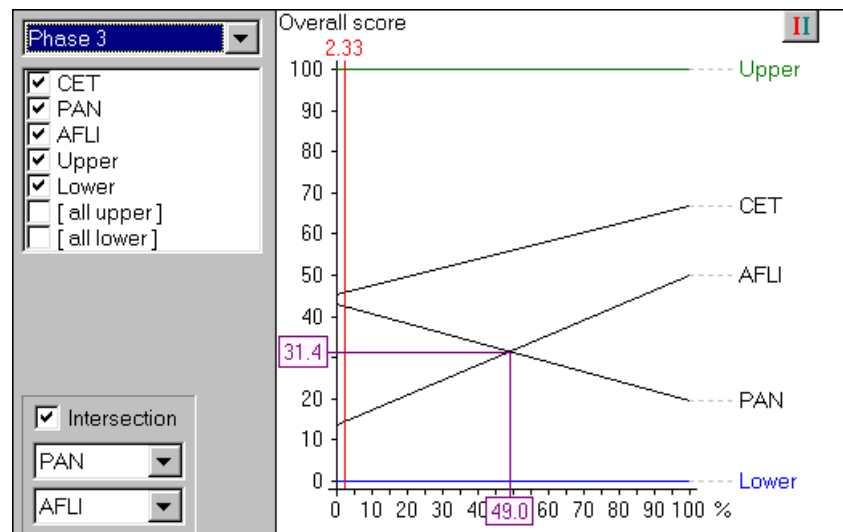
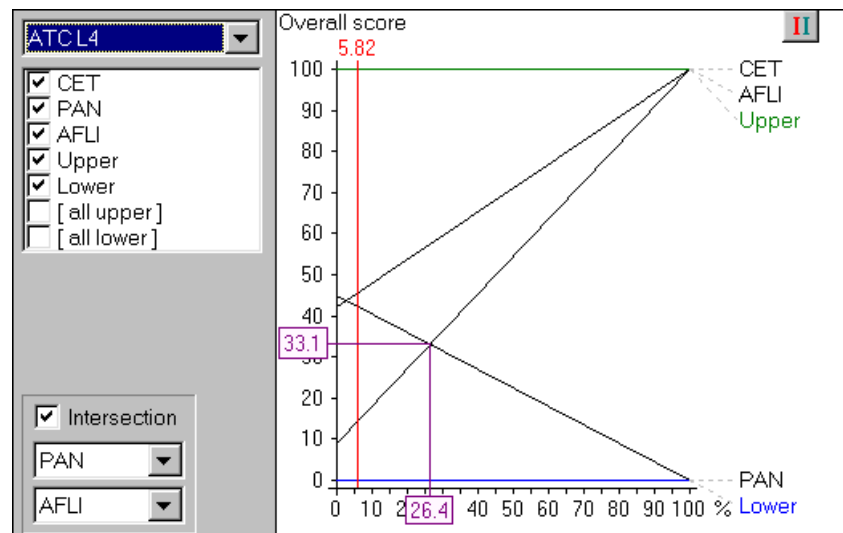


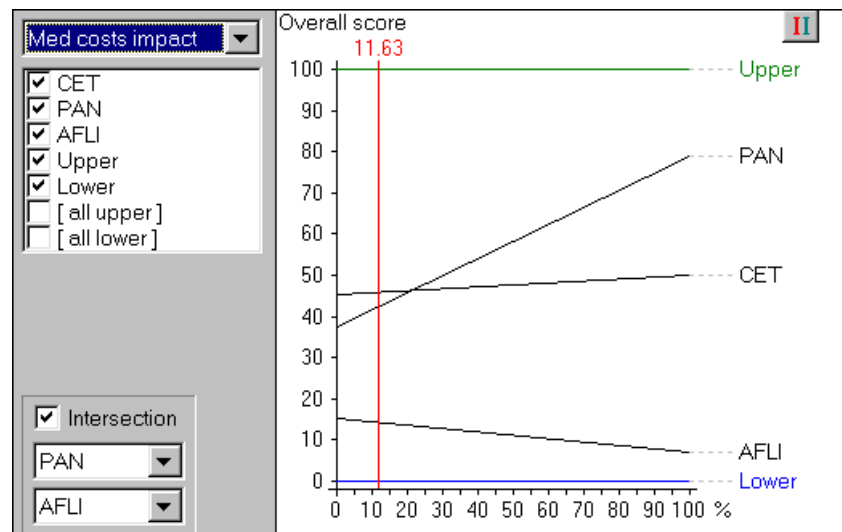
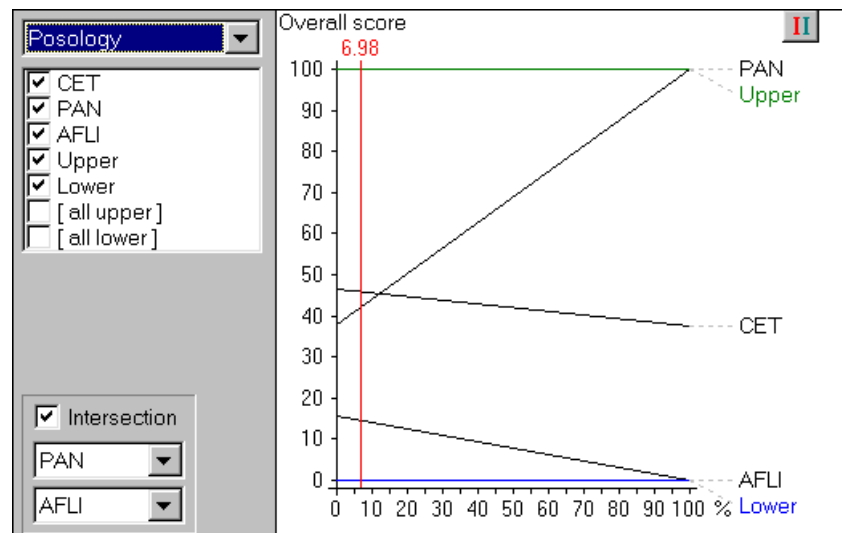


6.5.2 Sensitivity analysis on weights for Panitumumab vs Aflibercept plus FOLFIRI

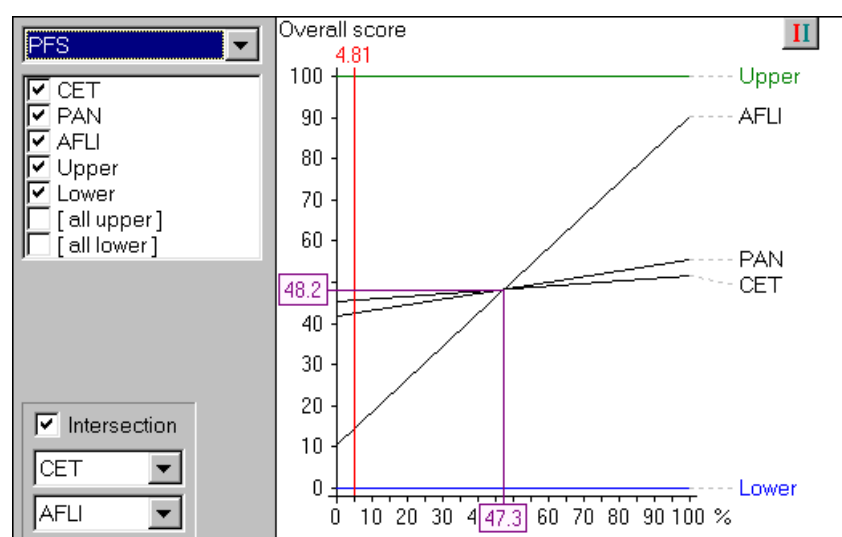
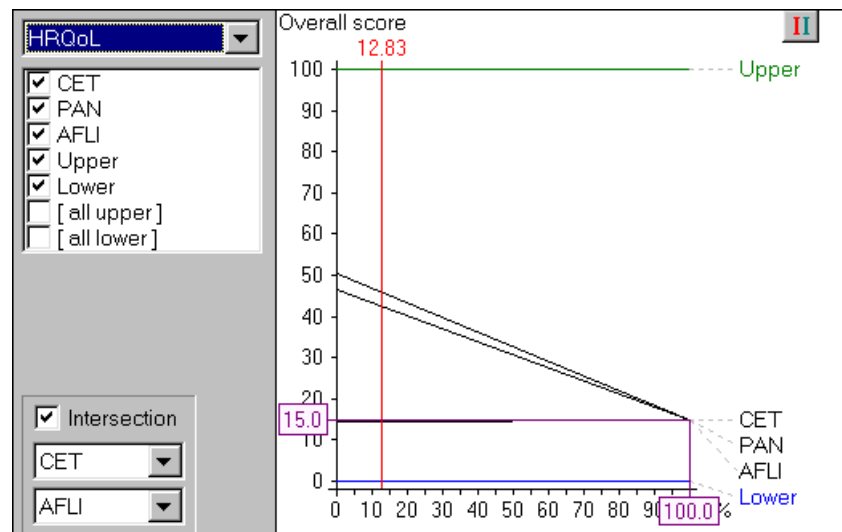
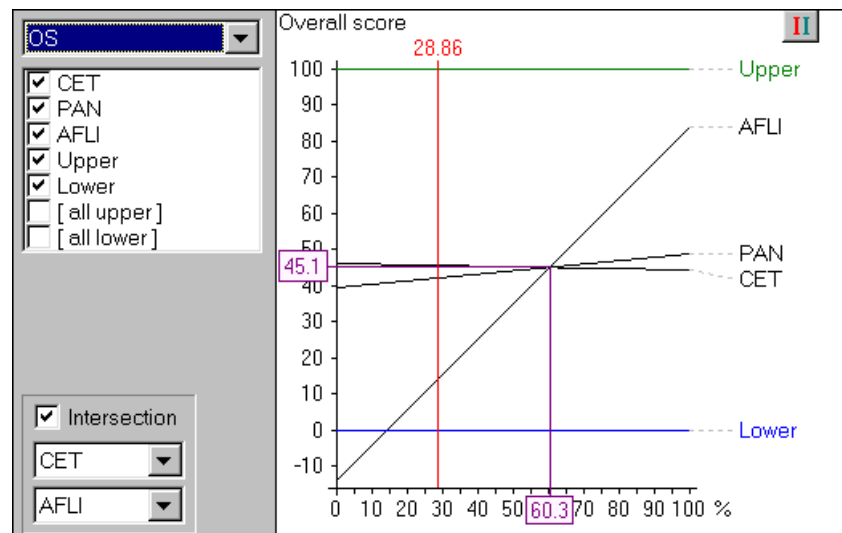


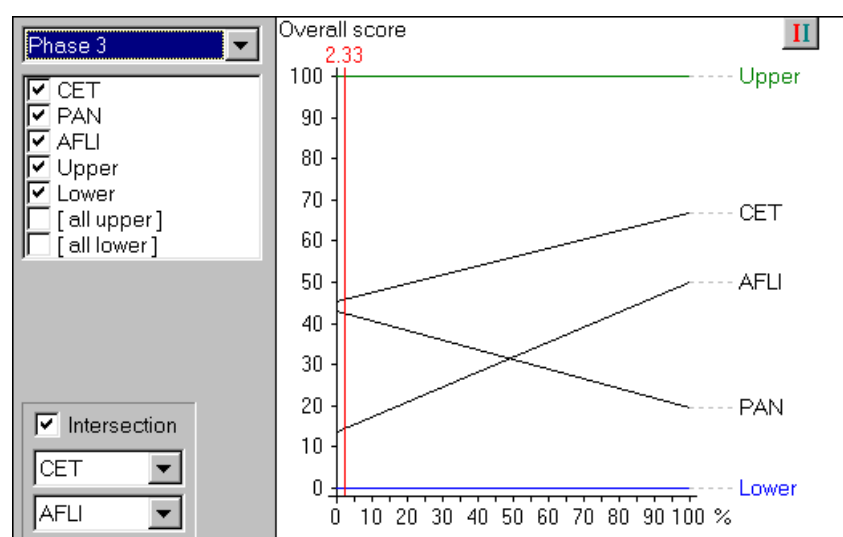
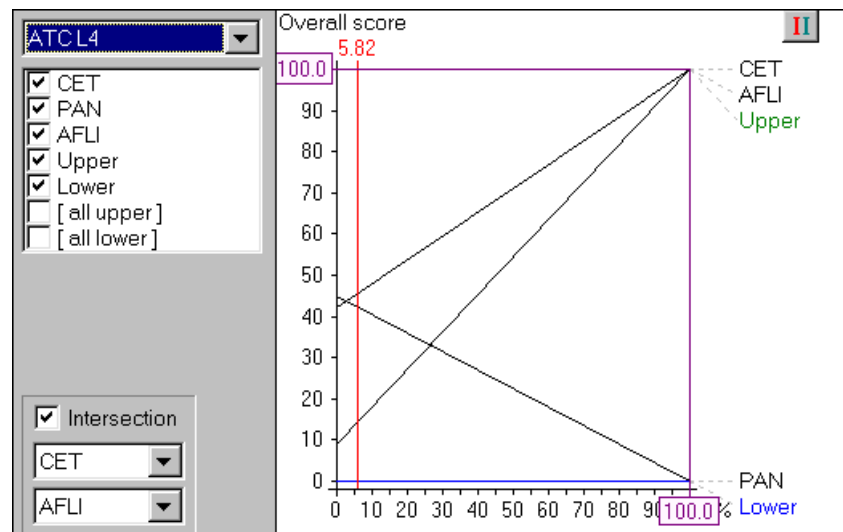
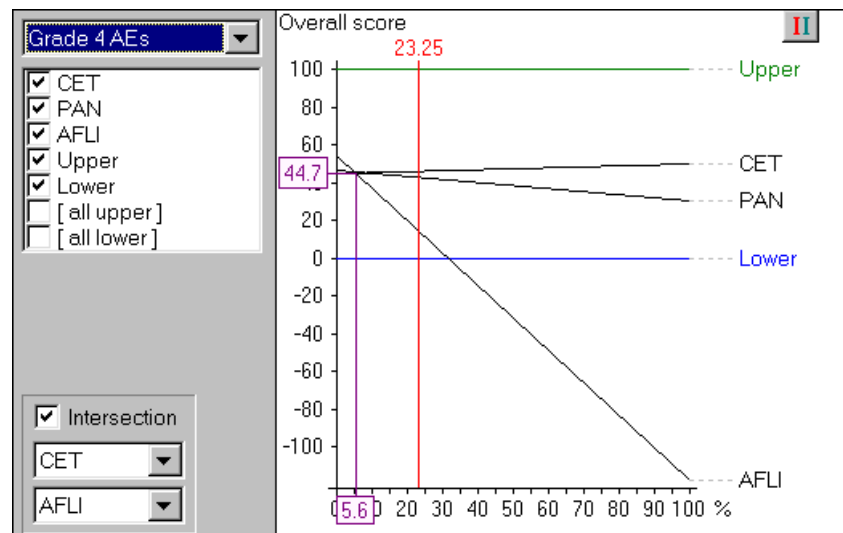


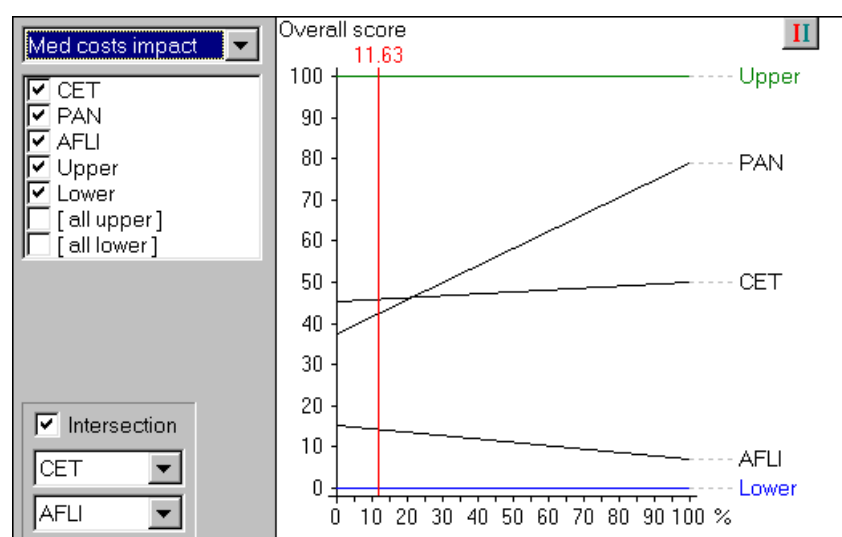
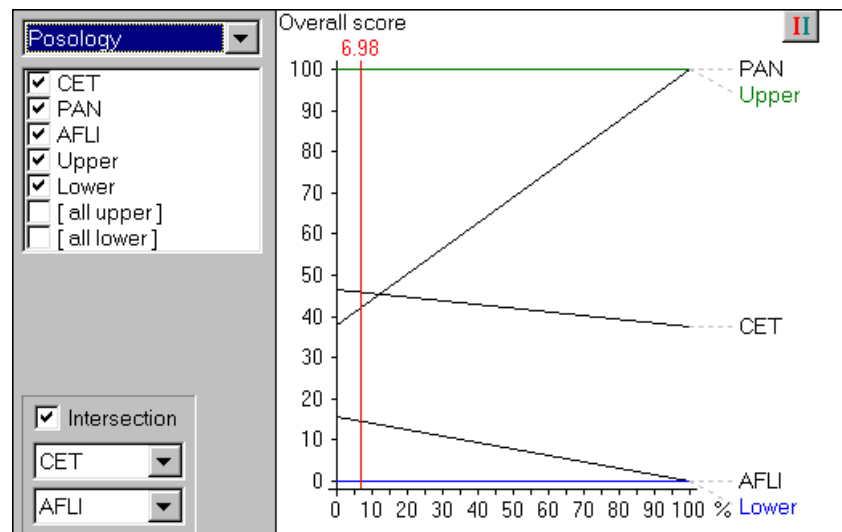
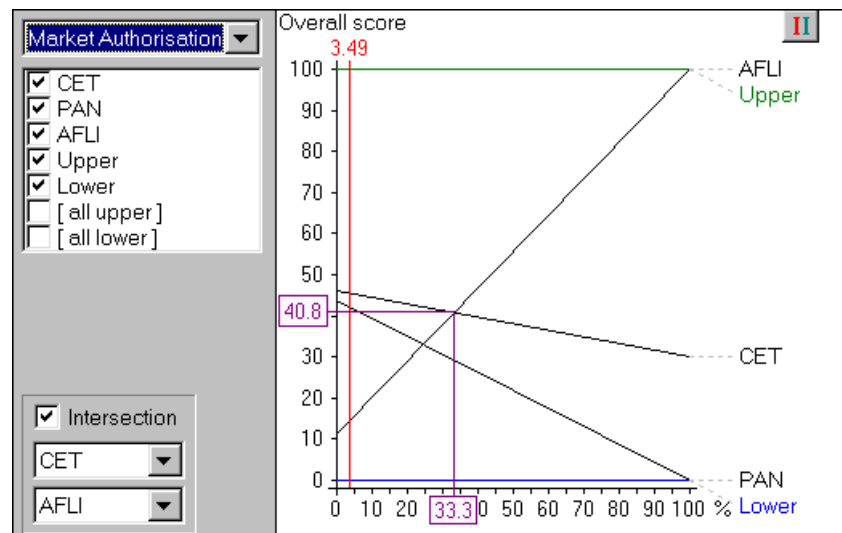





6.5.3 Sensitivity analysis on weights for Cetuximab vs Aflibercept plus FOLFIRI







6.5.4 Robustness analysis on reference levels:

	Upper	[all upper]	CET	PAN	AFLI	Lower	[all lower]
Upper	=	=	▲	▲	▲	▲	▲
[all upper]	=	=	▲	▲	▲	▲	▲
CET			=	+	+	▲	▲
PAN				=	+	▲	▲
AFLI					=	+	+
Lower						=	=
[all lower]						=	=

Local information			
	ordinal	MACBETH	cardinal
OS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±5%
HRQoL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±5%
PFS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±5%
Grade 4 AEs	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±5%
ATC L4	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>
Phase 3	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> ±5%
Market Authorisation	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> ±5%
Posology	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±5%
Med costs impact	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> ±5%

Global information		
ordinal	MACBETH	cardinal
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> ±0%

Diff

Red triangles denote “dominance” (an option dominates another if it is at least as attractive as the other in all criteria and it is more attractive than the other in at least one criterion). Green crosses denote “additive dominance” (an option additively dominates another if it is always found to be more attractive than the other through the use of an additive model under a set of information constraints).

For more information please see M-MACBETH user manual ⁴⁴⁸.

Appendix to Chapter 7

7.1 Methodological Framework

At first, as part of the problem structuring phase, the decision problem and the aims of the analysis are defined, and the relevant decision-makers and other key stakeholders are identified. Next, as part of the model-building phase, objectives and/or relevant criteria are identified in order to reflect decision-makers' goals and areas of concern, and attributes are selected to operationalise the criteria. In addition, under the same phase, selection of the alternative options takes place and evidence on their performance across the selected criteria is identified. Following that, under the model assessment phase, the performance of options against the criteria is assessed (i.e. scoring) and criteria are weighted according to their relative importance (i.e. weighing). Subsequently, as part of the appraisal phase, scores and weights are combined in order to produce overall WPV scores, taking the form of a value index (i.e. aggregation). In combination with sensitivity analysis, the results are examined and their robustness is determined. Finally, as part of action planning, the outcome of the analysis can be used to inform resource allocation decisions, of a coverage or pricing nature.

7.2 Evidence Considered and Alternative Treatments Compared

As part of NICE TA255 ⁴⁷², for the case of cabazitaxel in combination with prednisone, NICE primarily considered clinical evidence coming from one phase III, randomised, open-label, multicentre trial (*TROPIC*) investigating the use of cabazitaxel plus prednisone (or prednisolone) compared to mitoxantrone plus prednisone (or prednisolone) in men with hormone-refractory metastatic prostate cancer. Patients had to be aged over 18 years with an Eastern Cooperative Oncology Group (ECOG) performance score of 0–2, and with evidence of disease progression during or after completion of docetaxel-containing treatment ⁴⁶¹. The same clinical trial was used by TLV as part of a health economic exercise (no formal appraisal).

As part of TA259 ⁴⁶⁶, the decision problem considered whether treatment with abiraterone plus prednisolone was clinically effective compared with mitoxantrone (with or without prednisolone) or best supportive care for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen. NICE primarily considered clinical evidence coming from a phase III, placebo-controlled, randomised,

double-blind, multicentre trial (*COU-AA-301*), investigating the use of abiraterone in combination with prednisone (or prednisolone) versus placebo in combination with prednisone (or prednisolone), in men whose disease had progressed on or after docetaxel therapy ⁴⁶². Patients were aged over 18 years, with an Eastern Cooperative Oncology Group (ECOG) performance score of 0–2. A similar decision problem was adopted in TLV TA4774/2014 for the case of abiraterone in combination with prednisolone versus prednisolone on its own for patients who had received docetaxel or comparable chemotherapy, with clinical evidence coming from the *COU-AA-301* trial ⁴⁷⁴.

As part of TA316 ⁴⁶⁷, for the case of enzalutamide NICE primarily considered clinical evidence coming from a phase III randomised double-blind placebo-controlled study (*AFFIRM*) which investigated the use of enzalutamide plus best supportive care¹²⁸ (i.e. with or without the use of prednisone or other glucocorticoids) compared with placebo plus best supportive care ⁴⁶³. Eligible patients were aged over 18 years, with metastatic hormone-relapsed prostate cancer who had previously received 1 or 2 cytotoxic chemotherapy regimens, at least 1 of which contained docetaxel. Patients who had received abiraterone or treatment with any other investigational agents that block androgen synthesis were excluded. A similar decision problem was adopted in TLV TA2775/2013 for the case of enzalutamide versus best supportive care for patients who had progressed during or after docetaxel treatment, with clinical evidence base from the *AFFIRM* study ⁴⁷³.

In addition, as part of NICE TA316 evidence for abiraterone plus prednisone from the *COU-AA-301* trial was also considered in order to indirectly compare enzalutamide versus abiraterone (plus prednisone) using placebo as a common comparator whereas TLV TA4852/2014 used the same pivotal trials to compare enzalutamide versus abiraterone, either (1) when treatment with hormonal therapy has not worked or when treatment has not worked in men without symptoms or with only mild symptoms that do not need chemotherapy yet (i.e. pre-chemotherapy), or (2) adult men with progressive disease during or after docataxel-based chemotherapy (i.e. post-chemotherapy); none of these two scopes were considered.

¹²⁸ Best supportive care in *AFFIRM* could include radiopharmaceuticals, analgesics, bisphosphonates, hormonal therapies, corticosteroids, and radiotherapy

7.3 MCDA Technique

MACBETH uses seven semantic categories ranging between “no difference” to “extreme difference”, in order to distinguish between the value of different attribute levels. Based on these qualitative judgements of difference and, by analysing judgmental inconsistencies, it facilitates the move from ordinal preference modeling, a cognitively less demanding elicitation of preferences, to a quantitative value function. An example of the type of questioning being asked would be “What do you judge is the difference of value between x' and x'' ?” where x' and x'' are two different attribute levels of attribute x , across the plausible range (i.e. $x_* \leq x', x'' \leq x^*$). The approach has evolved through the course of theoretical research and real world practical applications, making it an interactive decision support system that facilitates decision-makers’ communication.

Following the elicitation of value functions, criteria baseline weights can be elicited. Questions of direct importance for a criterion such as “*How important is a given criterion?*” are known to be as one of the most common mistakes when making value trade-offs because they are assessing them independent of the respective attribute ranges ²⁰². In contrast, indirect weighting technique that assess value trade-offs in tandem with the respective ranges of attributes should be employed. For example, the quantitative swing weighting technique asks for judgments of relative value between ‘swings’ (i.e. changes) from standard lower level x_* to higher reference level x^* on each x -th attribute) taking the form “*How would you rank the relative importance of the criteria, considering their attributes ranges relative to 100 for the highest-ranked criterion considering its range?*”. Each swing, i.e. a relative change from a lower attribute level to a higher attribute level, is valued between 0 and 100, with the most valuable swing anchored as 100 ⁵⁴. Normalised weights are then calculated, as a proportion of each swing weight, so the normalised weights summed to 100%. Instead, relative attribute weights were calculated using an alternative qualitative swing weighting protocol, by using the MACBETH procedure to elicit the differences in attractiveness between the lower and higher reference levels of the different attributes, initially at individual level and then at criteria cluster level (i.e. by considering multiple attribute swings on the same time) ^{203,407}.

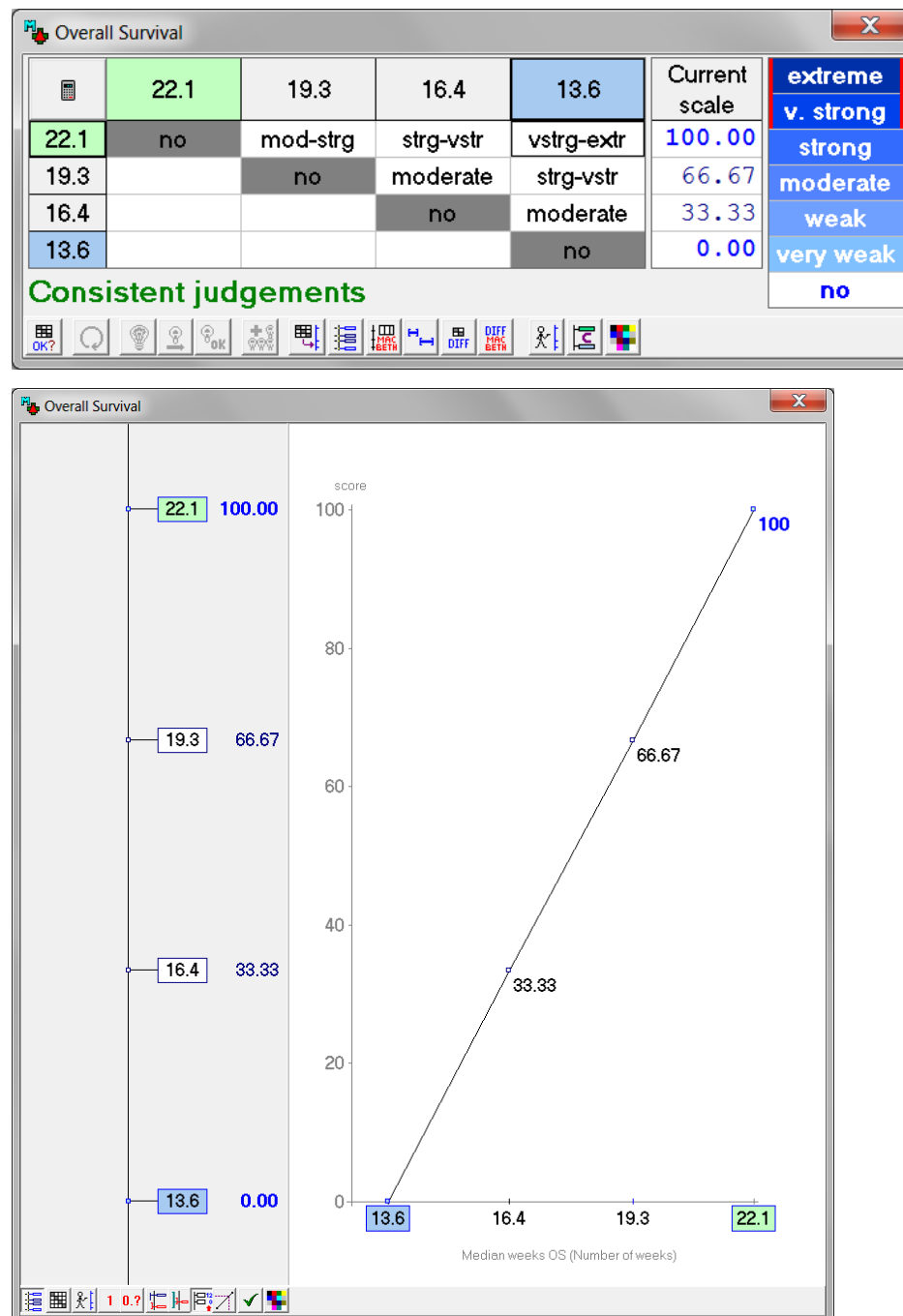
Finally criteria preference value scores and the respective weights can be combined together through an additive aggregation approach as described in Eq. X (if

the adequate conditions of complete and transitive preferences are met as well as multi-attribute preferential independence conditions – see ⁵⁴ and section 1.1.4).

7.4 Criteria Validation and Amended Value Tree for Metastatic Prostate Cancer¹²⁹

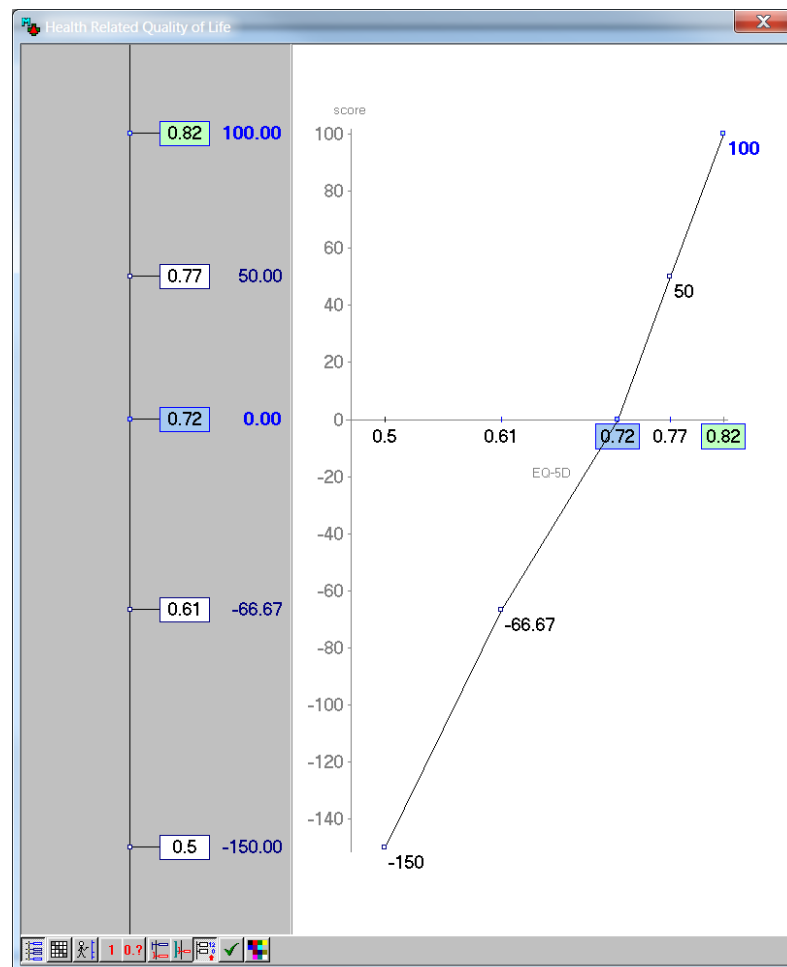
¹²⁹ images produced using the M-MACBETH (beta) software version 3.0.0

7.4.1 Example of value judgements matrix for the Overall Survival attribute and their conversion into value functions

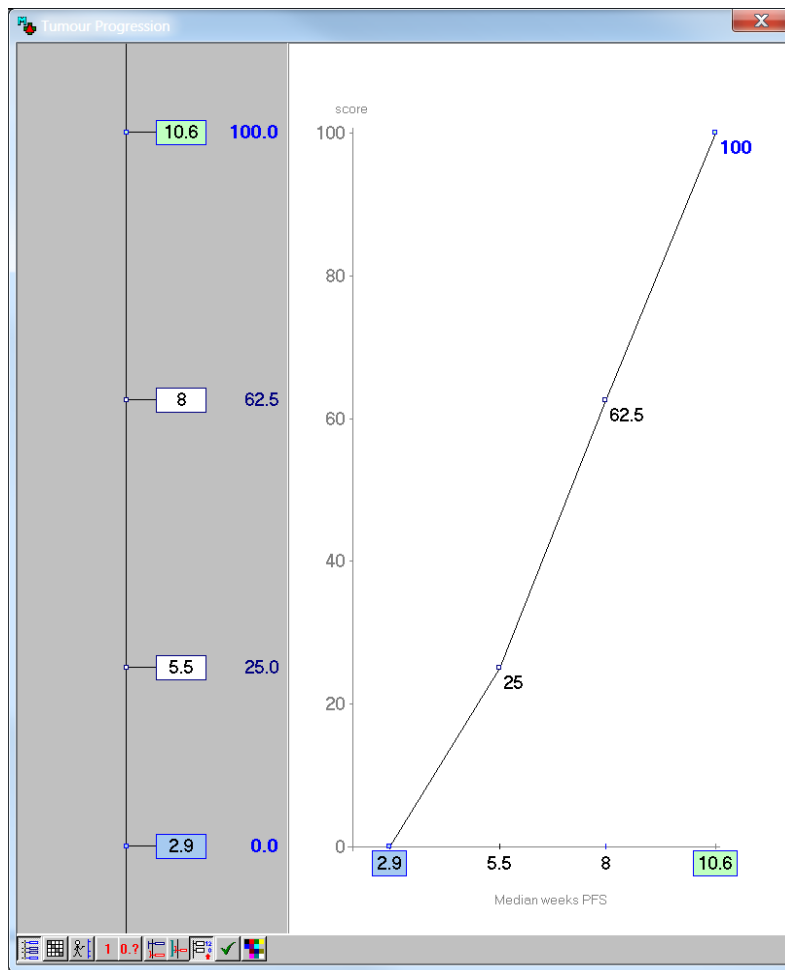


Caption: In the Overall Survival example, the question asked was the following: “What do you judge to be the difference of value between 13.6 months OS and 22.1 months OS? No difference, very weak, weak, moderate, strong, very strong, or extreme?” Once a consensus was reached, the next question came along: “What do you judge to be the difference of value between 16.4 months OS and 22.1 months OS? No difference, very weak, weak, moderate, strong, very strong, or extreme?” The same process was followed until value judgments for all the different combinations of attribute levels were elicited, filling in the different rows from the right-hand side (i.e. lower range) to the left-hand side (i.e. higher range).

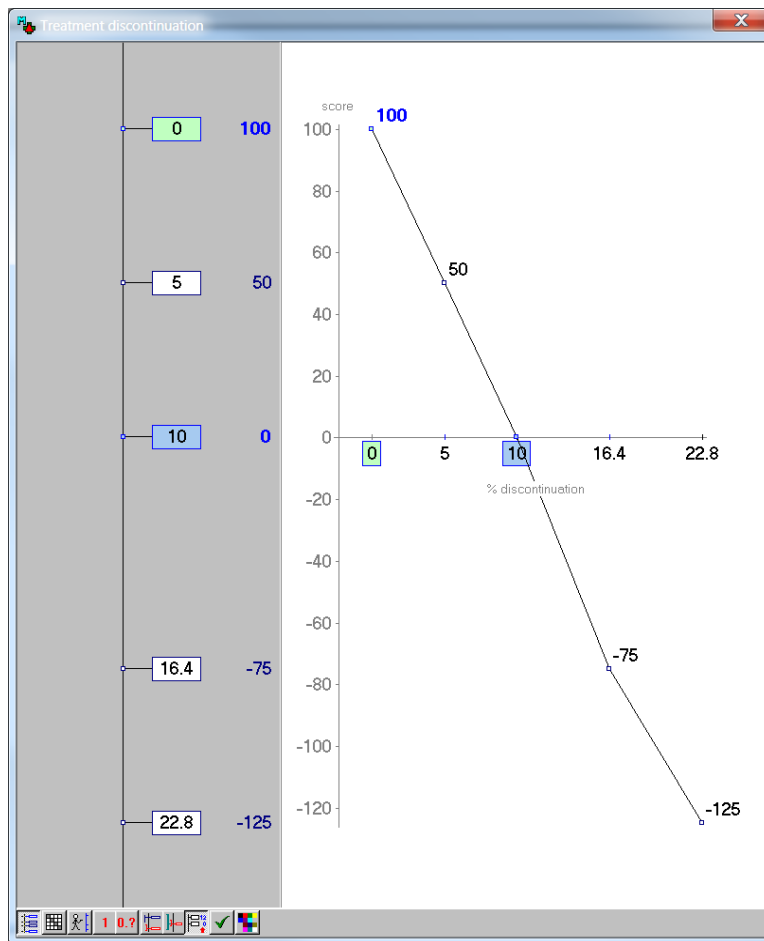
7.4.2 Value scale for the Health Related Quality of Life attribute



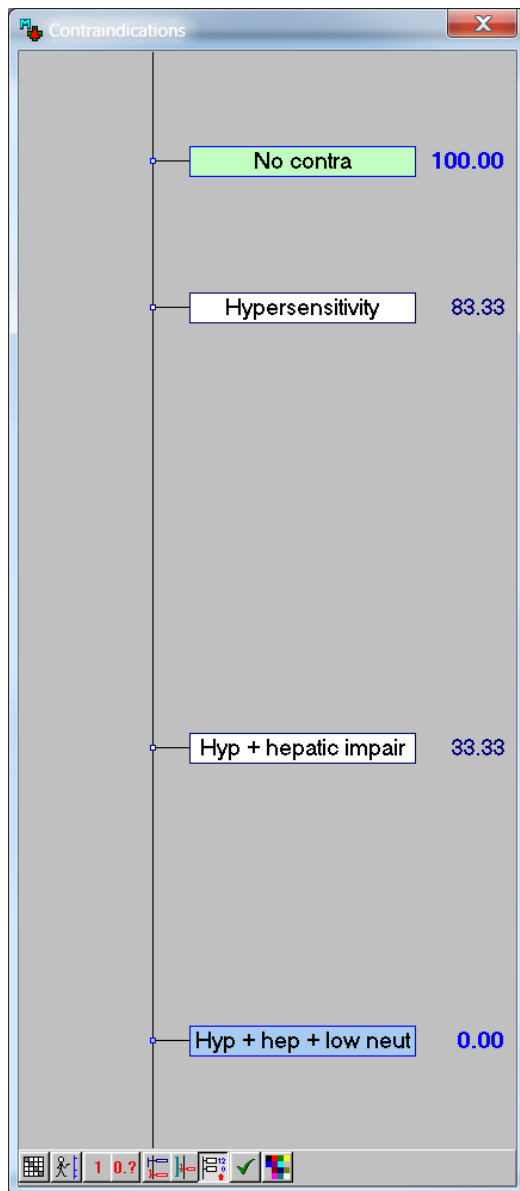
7.4.3 Value scale for the Tumour Progression attribute



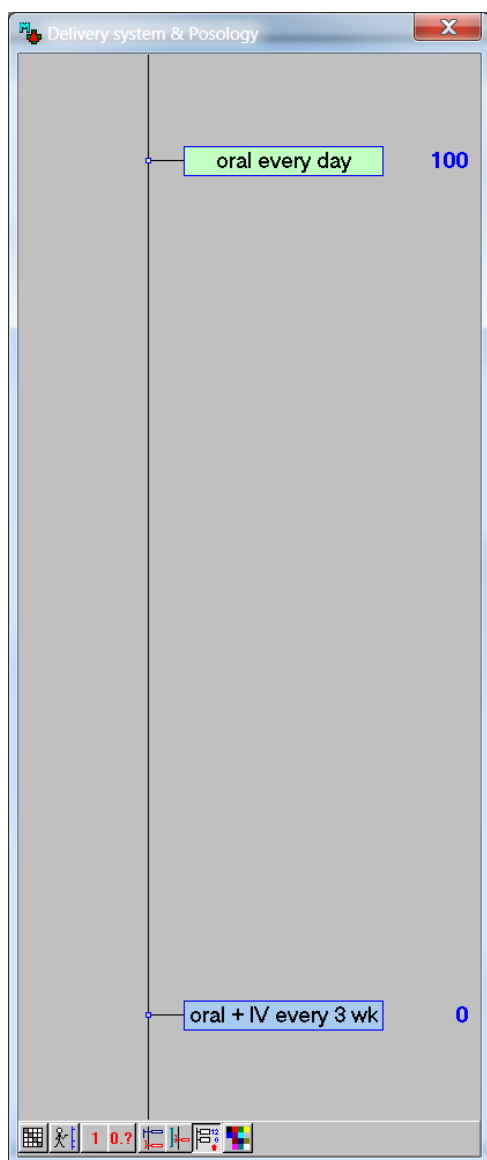
7.4.4 Value scale for the Treatment discontinuation attribute



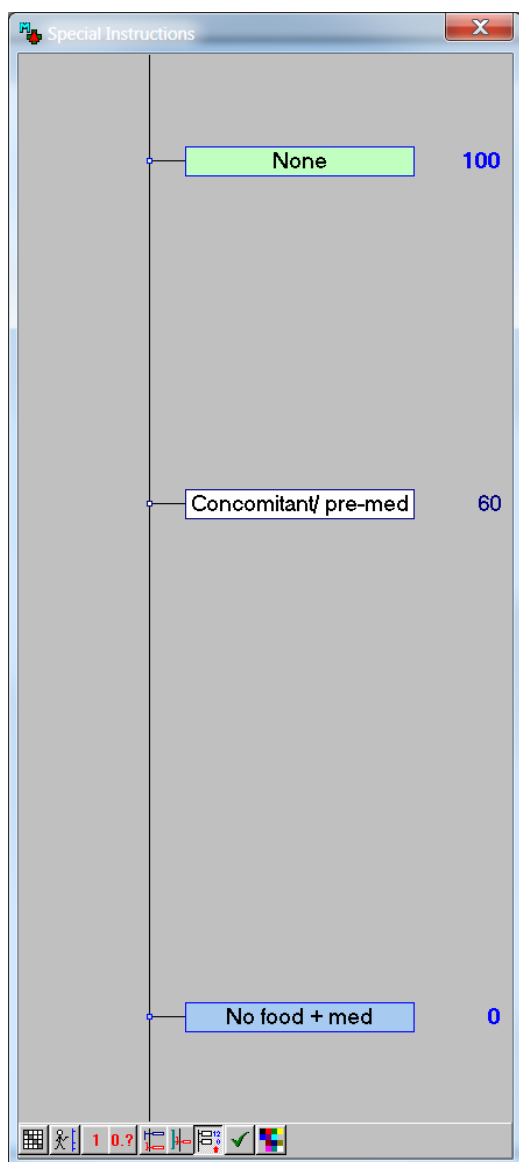
7.4.5 Value scale for the Contraindications attribute



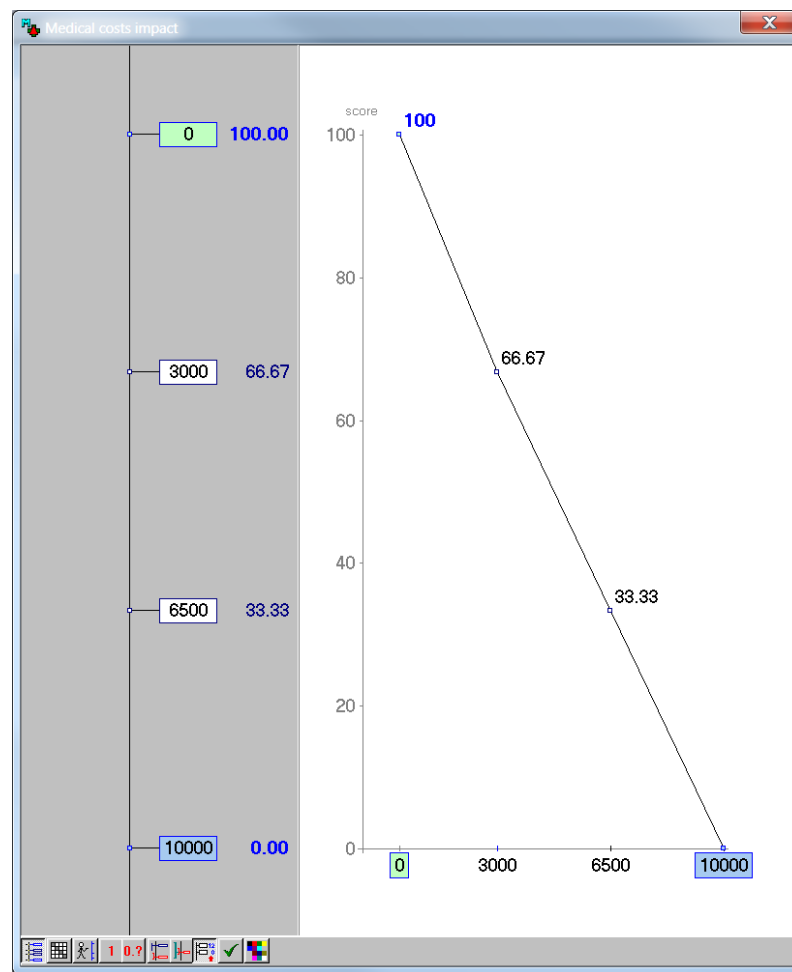
7.4.6 Value scale for the Delivery System & Posology attribute



7.4.6 Value scale for the Special Instructions attribute

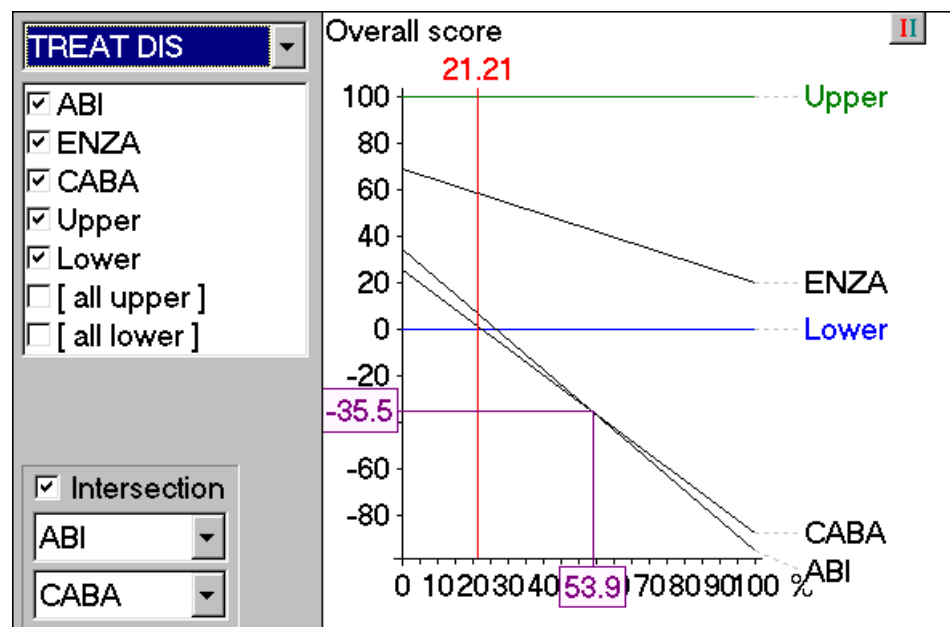
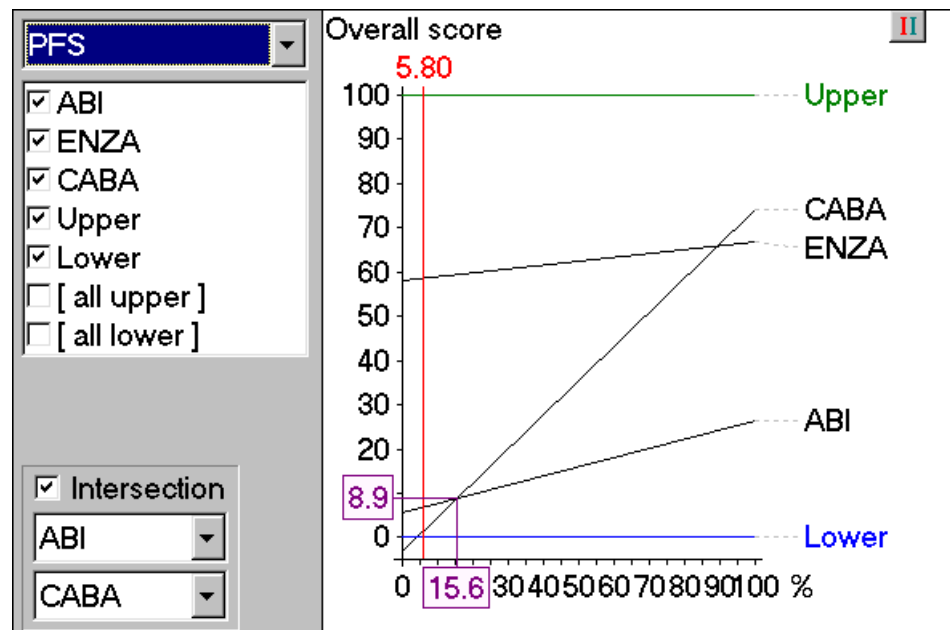


7.4.7 Value scale for the Medical Costs Impact attribute

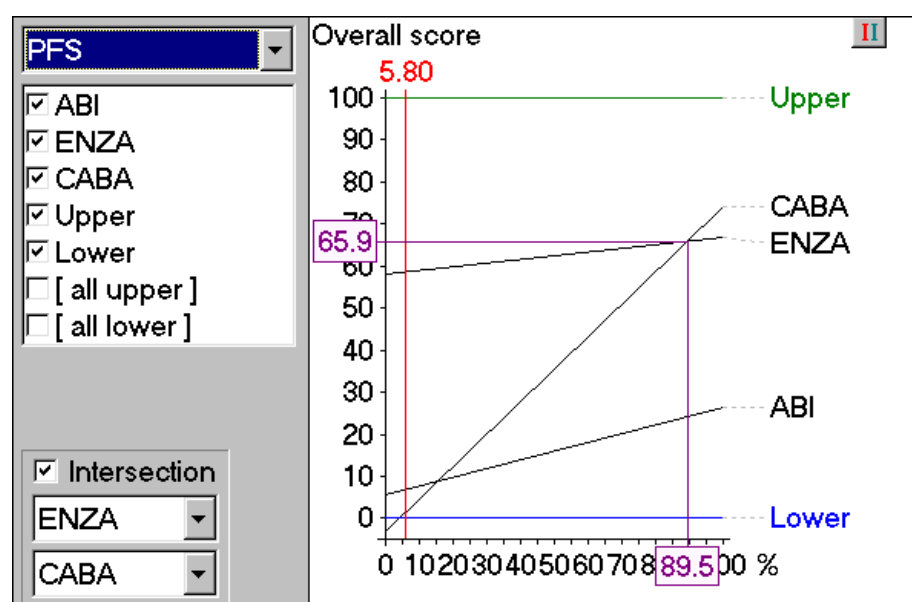
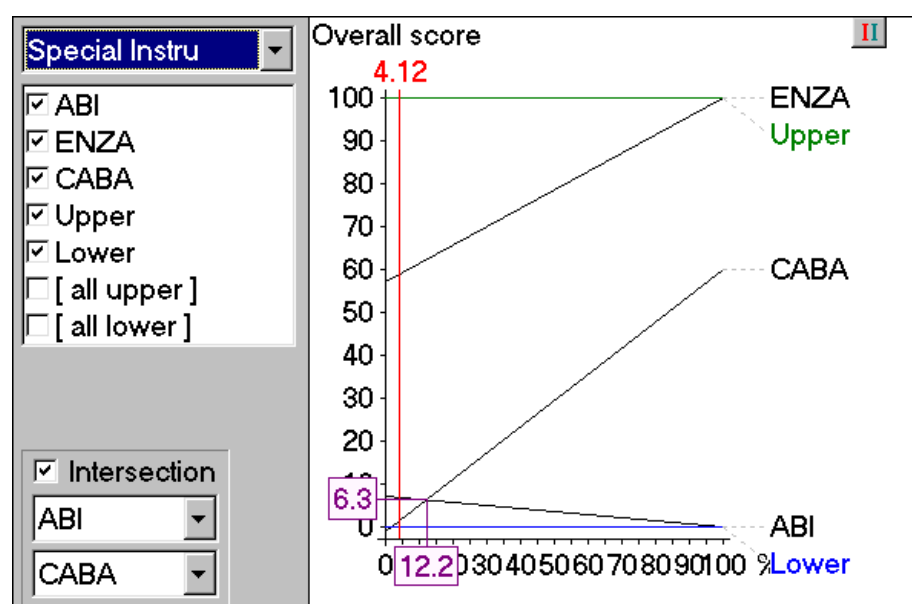


7.5 Sensitivity and Robustness Analysis¹³⁰

7.5.1 Exploring changes in the relative weights of Progression Free Survival (PFS), Treatment discontinuation (TREAT DIS), Special Instructions (Special Instru) and their impact on treatments' rankings



¹³⁰ images produced using the M-MACBETH (beta) software version 3.0.0



7.5.2 Exploring changes in the ranges of the reference levels and their impact on treatment rankings.

☰	Upper	[all upper]	ENZA	ABI	Lower	[all lower]	CABA
Upper	=	=	▲	▲	▲	▲	▲
[all upper]	=	=	▲	▲	▲	▲	▲
ENZA			=	▲	▲	▲	+
ABI				=	?	?	+
Lower				?	=	=	?
[all lower]				?	=	=	?
CABA					?	?	=

Local information			
	ordinal	MACBETH	cardinal
OS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
HRQoL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
PFS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
TREAT DIS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
Contra	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
Delivery Posology	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Special Instru	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%
Medical costs	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> ±10%

Global information		
ordinal	MACBETH	cardinal
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/> ±0%

Red triangles denote “dominance” (an option dominates another if it is at least as attractive as the other in all criteria and it is more attractive than the other in at least one criterion). Green crosses denote “additive dominance” (an option additively dominates another if it is always found to be more attractive than the other through the use of an additive model under a set of information constraints).

For more information please see M-MACBETH user manual available via: <http://www.m-macbeth.com/help/pdf/M-MACBETH%202.4.0%20Users%20Guide.pdf>

Appendix to Published Articles

- a) Angelis, Aris & Kanavos, Panos, 2016. Value-based assessment of new medical technologies: towards a robust methodological framework for the application of multiple criteria decision analysis in the context of health technology assessment. *Pharmacoeconomics* 34 (5), 435-446
Available via: <https://link.springer.com/article/10.1007/s40273-015-0370-z>

- b) Angelis, Aris, Lange, Ansgar & Kanavos, Panos, 2018. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *The European Journal of Health Economics* 19: 123
Available via: <https://link.springer.com/article/10.1007/s10198-017-0871-0>

- c) Angelis, Aris & Kanavos, Panos, 2017. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: the Advance Value Framework. *Social Science & Medicine* 188, 137-156
Available via:
<https://www.sciencedirect.com/science/article/pii/S0277953617303933>

- d) Angelis, Aris & Kanavos, Panos, 2016. Critique of the American Society of Clinical Oncology value assessment framework for cancer treatments: putting methodologic robustness first. *Journal of Clinical Oncology* 34 (24), 2935-2936
Available via:
<http://ascopubs.org/doi/abs/10.1200/JCO.2015.64.9673>

- e) Angelis, A., Kanavos, P. & Montibeller, G., 2017. Resource Allocation and Priority Setting in Health Care: A Multi-criteria Decision Analysis Problem of Value? *Global Policy* 8(S2), pp.76–83.
Available via: <https://onlinelibrary.wiley.com/doi/full/10.1111/1758-5899.12387>

- f) Tervonen, Tommi et al., 2015. Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention. *Medical Decision Making* 35 (7), 859-871.

Available via:

<http://journals.sagepub.com/doi/abs/10.1177/0272989X15587005>

